

DIGITAL HOLOGRAPHY AND OPTICAL CONTOURING

YAN LI

**A thesis submitted in partial fulfilment of the
requirements of Liverpool John Moores University
for the degree of Doctor of Philosophy**

January 2009

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge the supervision of Prof. Michael Lalor and Prof. David Burton during this work. I thank Dr. Francis Lilley for the technical support on the equipment selections and helpful corrections of this thesis. I cherished the discussions with all of them to make each progress of my project. Other colleagues in Coherent & Electro-Optics Research Group also provided valuable help, especially Dr. Hussein Abdul-Rahman who is always patient and helpful to answer my questions about programming and phase unwrapping problems.

I am grateful to our secretaries, Mrs. Margaret Sutton and Miss Helen Pottle. They helped me a lot in many administration issues so that I could concentrate on my research.

Thanks also go to GERI for the financial support of these years. Without it, I was not able to be here to do this project and this thesis could not be finished.

I am forever indebted to my parents for their understanding, endless patience and encouragement when it was most required. There were also many dear friends who showed their support to me with words of encouragement, nice meals and prayers.

Finally, I would like to give my special thanks to Lord Jesus Christ of whose love I do not deserve. Thank Him for the great grace and perseverance He granted me during the course of this project. It has never been an easy task to achieve a PhD degree. However He gave me more than that.

ABSTRACT

Digital Holography and Optical Contouring

Digital holography is a technique for the recording of holograms via CCD/CMOS devices and enables their subsequent numerical reconstruction within computers, thus avoiding the photographic processes that are used in optical holography. This thesis investigates the various techniques which have been developed for digital holography. It develops and successfully demonstrates a number of refinements and additions in order to enhance the performance of the method and extend its applicability. The thesis contributes to both the experimental and numerical analysis aspects of digital holography.

Regarding experimental work: the thesis includes a comprehensive review and critique of the experimental arrangements used by other workers and actually implements and investigates a number of these in order to compare performance. Enhancements to these existing methods are proposed, and new methods developed, aimed at addressing some of the perceived short-comings of the method. Regarding the experimental aspects, the thesis specifically develops:

- Super-resolution methods, introduced in order to restore the spatial frequencies that are lost or degraded during the hologram recording process, a problem which is caused by the limited resolution of CCD/CMOS devices.
- Arrangements for combating problems in digital holography such as: dominance of the zero order term, the twin image problem and excessive speckle noise.
- Fibre-based systems linked to tunable lasers, including a comprehensive analysis of the effects of: signal attenuation, noise and laser instability within such systems.
- Two-source arrangements for contouring, including investigating the limitations on achievable accuracy with such systems.

Regarding the numerical processing, the thesis focuses on three main areas. Firstly, the numerical calculation of the Fresnel-Kirchhoff integral, which is of vital importance in performing the numerical reconstruction of digital holograms. The Fresnel approximation and the convolution approach are the two most common methods used to perform numerical reconstruction. The results produced by these two methods for both simulated holograms and real holograms, created using our experimental systems, are presented and discussed.

Secondly, the problems of the zero order term, twin image and speckle noise are tackled from a numerical processing point of view, complementing the experimental attack on these problems. A digital filtering method is proposed for use with reflective macroscopic objects, in order to suppress both the zero-order term and the twin image.

Thirdly, for the two-source contouring technique, the following issues have been discussed and thoroughly analysed: the effects of the linear factor, the use of noise reduction filters, different phase unwrapping algorithms, the application of the super-resolution method, and errors in the illumination angle. Practical 3D measurement of a real object, of known geometry, is used as a benchmark for the accuracy improvements achievable via the use of these digital signal processing techniques within the numerical reconstruction stage.

The thesis closes by seeking to draw practical conclusions from both the experimental and numerical aspects of the investigation, which it is hoped will be of value to those aiming to use digital holography as a metrology tool.

Contents

Acknowledgements.....	i
Abstract.....	ii
Abbreviations and symbols.....	vi
1 Introduction.....	1
1.1 An overview of 3D shape measurement using optical methods.....	1
1.2 The conception and principle of holography.....	4
1.3 The development of optical holography.....	7
1.4 The development of digital holography.....	11
1.5 Some problems in digital holography.....	13
1.5.1 The spatial frequency requirements	13
1.5.2 The algorithm for numerical reconstruction	14
1.5.3 The suppression of the zero-order term and twin image.....	14
1.6 Objective of this thesis	14
1.7 Contributions of this thesis.....	15
1.8 The structure of this thesis.....	15
References.....	17
2 Numerical Reconstruction in Digital Holography	22
2.1 General principle of digital holography.....	22
2.2 Numerical reconstruction of digital holograms	27
2.2.1 Reconstruction by the Fresnel approximation	27
2.2.2 Reconstruction by the convolution approach.....	38
2.2.3 Comparison of Fresnel approximation and convolution approach	45
2.2.4 Lensless digital holography (digital Fourier holography).....	47
2.2.5 Angular spectrum method.....	49
2.2.6 Phase retrieval methods.....	50
2.2.7 Other algorithms for numerical reconstruction	52

2.3	Conclusions	52
	References.....	54
3	Suppression of The Zero-Order and Twin-Image Terms	56
3.1	Introduction	56
3.2	Suppression of zero-order term and twin image in off-axis holograms	57
3.2.1	Methods based on digital image processing.....	58
3.2.2	Methods based on experiments	72
3.3	Zero-order and twin image suppression for in-line holograms	81
3.4	Conclusions	88
	References.....	90
4	Enhancements of Digital Holography	91
4.1	Introduction	91
4.2	Recording conditions.....	91
4.2.1	Typical configurations in digital holographic systems.....	91
4.2.2	The intensity of the reference wave	93
4.3	Reduction of speckle noise	97
4.3.1	Digital image processing.....	97
4.3.2	Multiple holograms	100
4.3.3	Deconvolution of the aperture function	106
4.4	Resolution improvement.....	108
4.5	Conclusions	121
	References.....	123
5	Optical contouring of 3D objects	125
5.1	Introduction	125
5.2	Contour generation	126
5.3	Two-source methods.....	127
5.3.1	The conventional two-source method	127
5.3.2	The two-source method in digital holography	128

5.3.3	Experimental system for the two-source method.....	130
5.3.4	Experimental results for the two-source method	132
5.4	Two-wavelength methods.....	150
5.4.1	The conventional two-wavelength method	150
5.4.2	The two-wavelength method in digital holography	152
5.4.3	Experimental results when using the two-wavelength method.....	155
5.5	Two-refractive-index methods.....	158
5.5.1	The conventional two-refractive-index method.....	158
5.5.2	The two-refractive-index method in digital holography	159
5.6	Decorrelation effects in optical contouring	161
5.7	Conclusions	162
	References.....	165
6	Conclusions and Future Work	167
6.1	Conclusions	167
6.2	Future work.....	168
	References.....	174
	Bibliography.....	175
	Appendix.....	176

ABBREVIATIONS AND SYMBOLS

3D	Three dimensional
CCD	Charge coupled device
U_o	Complex amplitude of the object wave
U_r	Complex amplitude of the reference wave
E_o	Real amplitude of the object wave
E_r	Real amplitude of the reference wave
φ_o	Phase of the object wave
φ_r	Phase of the reference wave
U_h	Digital hologram
β	Constant dependent on the exposure characteristics of the hologram recording media
τ	Exposure time of recording the hologram
FFT	Fast Fourier transform
θ_{max}	Maximum angle between the object wave and the reference wave
f_{max}	Maximum spatial frequency that can be resolved
λ	Wavelength
Δx	Distance between neighbouring pixels of a CCD/CMOS camera in x axis
Δy	Distance between neighbouring pixels of a CCD/CMOS camera in y axis
SR	Super resolution
(ξ, η)	Point in the object plane
(x, y)	Point in the hologram plane
(ξ', η')	Point in the image plane
Γ	Diffacted field
ρ'	Distance between a point in the hologram plane and a point in the reconstruction plane
ρ	Distance between a point in the object plane and a point in the hologram plane
Q	Inclination factor
d	Recording distance between the object plane and the hologram plane
d'	Reconstruction distance between the reconstruction plane and the hologram plane
I	Intensity
φ	Phase
Re	Real part

Im	Imaginary part
$*$	Conjugate function
f	Focal length of a lens
$L(x, y)$	Lens function in the hologram plane
$P(\xi', \eta')$	Phase correction factor when introduced a lens
$\Delta\xi$	Distance between neighbouring pixels in x axis of the reconstruction plane
$\Delta\eta$	Distance between neighbouring pixels in y axis of the reconstruction plane
\otimes	Convolution operation

1 Introduction

Optical methods have been used as metrological tools for a long time. They are non-contact, non-destructive and can deliver high accuracy. With the combination of computers and other electronic devices, they have become faster, more reliable, more convenient and more robust. Amongst all the various applications, surface contouring is one of great interest. In industry, there is a need for accurate measurement of three-dimensional (3D) shape of objects to speed up product development and ensure manufacturing quality. Meanwhile in the medical field, knowledge of the 3D shape and other biological features of the suspected objects are helping doctors to make correct diagnoses, choose appropriate treatments and to effectively evaluate such treatments.

In this chapter, an overview of 3D shape measurement is initially presented. Then the conception and the development of holography as important tools to fulfil the task of 3D shape measurement are presented. The disadvantages and limitations of this technology are introduced later in the thesis. The objective, contribution and structure of the thesis are described in the last section of this chapter.

1.1 An overview of 3D shape measurement using optical methods

Various optical techniques have been developed for measuring 3D shape from a single position (Tiziani, 1997, Chen et al., 2000). These are summarized in Fig. 1.1. As the name suggests, the time-of-flight technique measures the flight time of a pulse from a laser or other light source to an object and back. The flight time of two pulses, one of which is reflected from the surface of the object to the receiving sensor and another passing through an optical fibre, are measured and compared to determine the distance. The typical resolution of this technique is approximately 1mm . Although with improvements in the equipment, the resolution can be better than $30\mu\text{m}$ at a stand off distance of 1m . Another similar technique is called light-in-flight holography. Either short temporal coherence light or a very short light pulse is used to generate a motion image of a propagating optical wavefront. The depth resolution may reach $6.5\mu\text{m}$ (Nilsson and Carlsson, 1998).

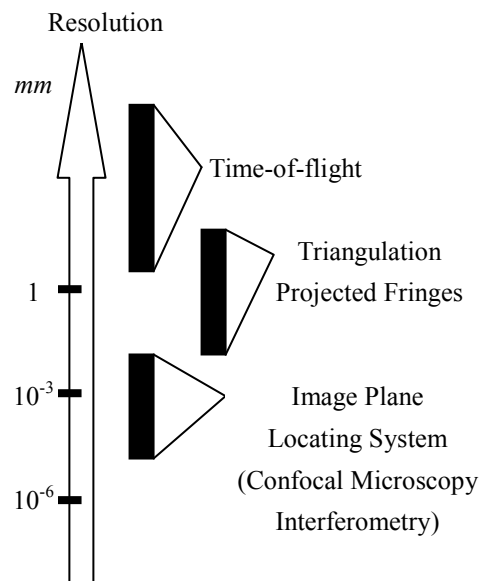


Fig. 1.1 Optical methods for 3-D measurements (redrawn from (Tiziani, 1997))

The active triangulation technique is a powerful tool for 3D shape measurement. A laser spot is projected on the surface of the object and is imaged via a Charge Coupled Device (CCD) or a Position Sensitive Detector (PSD). The lateral displacement of the spot image is related to the depth of the object. Its typical measurement range is 5mm to 250mm , and accuracy is about 1 part in 10,000, with a measurement frequency of 40kHz or higher (Ji and Leu, 1989, Keferstein and Marxer, 1998). Sensors based on the triangulation principle are suitable for inspection and measurement in an industrial environment. However both laser speckle and differences in the surface characteristics of the measured object from that of the calibration surface, limit the accuracy of the measurements.

Structured light techniques may also be categorized amongst the active triangulation methods, as they project either coded light or sinusoidal fringes onto the object. The height distribution of the surface of the object is encoded into a deformed fringe pattern, which is captured by an image acquisition sensor (Srinivasan et al., 1984). Then shape is directly decoded from this recorded image according to the calibrated system parameters. This technique is easy to implement in an industrial environment. By use of a computer controlled projector and camera, no moving parts have to be used and the full field measurement can be done within less than a second. In an optimized structured light projection system, the measurement accuracy may achieve 1 part in 20,000. However, the problem of shading,

which is inherent to all triangulation techniques, is also a limitation of structured light techniques.

Moiré can be traced back over many years as a measurement technique. The key to this technique is the moiré effect. When two sinusoidal gratings, i.e. a master grating and a reference grating, are superposed with a small angle between the grating lines, a fringe pattern of much lower frequency than the individual gratings can be generated and resolved by a CCD camera. It is often termed as mechanical interference, because the mathematical description of a Moiré pattern is the same as for the interference pattern formed by electromagnetic waves. Rapid image acquisition and phase shifting techniques have been introduced to Moiré systems to overcome environmental perturbations and to analyze the fringe pattern produced (Kujawinska et al., 1991, van Haasteren and Frankena, 1994). The typical measurement range of the phase shifting Moiré method is 1mm to 500mm with a resolution similar to that of structured light projection at $1/10$ to $1/100$ of a fringe.

Laser interferometry is probably one of the most commonly used techniques in metrology for high-resolution and high-precision measurement. The idea behind interferometric shape measurement is that the fringes are formed by variation of the sensitivity matrix that relates the geometric shape of an object to the measured optical phase. The matrix contains four variables, wavelength, refractive index, and illumination and observation directions. From these, three methods are derived. These are the dual or multiple wavelength methods (Haines and Hildebrand, 1965), the refractive index change methods (Zelenka and Varner, 1969), and the illumination direction variation/two sources methods (Abramson, 1976). The resolution of the two-wavelength method depends on the equivalent wavelength (Λ) and the phase resolution ($\Lambda/200$) which will be discussed later in this thesis. For example, two wavelengths of a tunable laser system ($0.635\mu\text{m}$ and $0.633\mu\text{m}$) will provide an equivalent wavelength of $201\mu\text{m}$ and a resolution of about $1\mu\text{m}$. The resolutions of two-refractive-index method and two-illumination-source method also depend upon the parameters used in both methods. This will be discussed in detail in Chapter 5. Another interferometric measurement technique is heterodyne interferometry using a frequency shift. The relative phase increases linearly with time and is measured electronically at the beat frequency of the reconstructed wavefields. It offers a high spatial resolution and interpolation up to $1/1000$ of a fringe. However, it usually requires more sophisticated electronic equipment and mechanical devices to scan the fringe pattern.

There are also other methods for optical 3-D shape measurement, such as laser speckle pattern sectioning (Dresel et al., 1992), photogrammetry, laser tracking systems and confocal microscopy (Wilson, 1990). The interested readers can find relevant articles in the bibliography.

Different methods have been developed to fulfil the requirements of various applications. For optically rough surfaces, the Moiré and fringe projection methods are widely used as they are less sensitive to environmental effects and can perform fast full field measurements. Also their measurement range is much larger than that of interferometry. However, if the size of the object is smaller than a few centimetres and the high accuracy is needed, interferometry is a more appropriate option. It was a barrier for holography that there was a necessary intermediate photographic recording set-up and wet-chemical processing stage. But with the advent of digital holography, holography can be performed nearly in real time and there is no longer any requirement for this time-consuming photographic process involved.

Before specifically describing the application of digital holography to the measurement of the surfaces of real objects, the latter part of this chapter will firstly present the background of this relatively young technology.

1.2 The conception and principle of holography

The word *holography* is derived from the Greek words, ‘holos’ which means ‘whole’ or ‘entire’ and ‘graphein’ which means ‘to write’. Holography is a method which records and reconstructs not only the irradiance of each point in an image, but also the direction in which the wave is propagating at that point. That is what its name means, i.e. a whole record, as compared to photography which only records the intensity of the optical wave at each point. Before a historical review of holography, it is helpful to know how to record an object by use of holographic technology. People often confuse the concepts of photography and holography, however, if the different ways that photography and holography work are correctly understood, the fundamental difference between them can be easily discerned.

Photography is a process that enables us to record and retain almost any scene perceived by the human eyes, which means it records the two dimensional irradiance distribution of an image. Each scene includes many reflecting or radiating points of light. The waves from

these elementary points all contribute to the object wave. Through the optical lens in the camera, this complex object wave is transformed into an image of the object and recorded on the photographic emulsion or other imaging devices.

Holography is quite different. It does not record an image of the object wave, but the object wave itself. Since all recording media, for example photographic emulsions or photonic detectors, only respond to irradiance, it is necessary to convert the phase information of the object wave into variations of irradiance. This makes holography an essentially different process from that of photography.

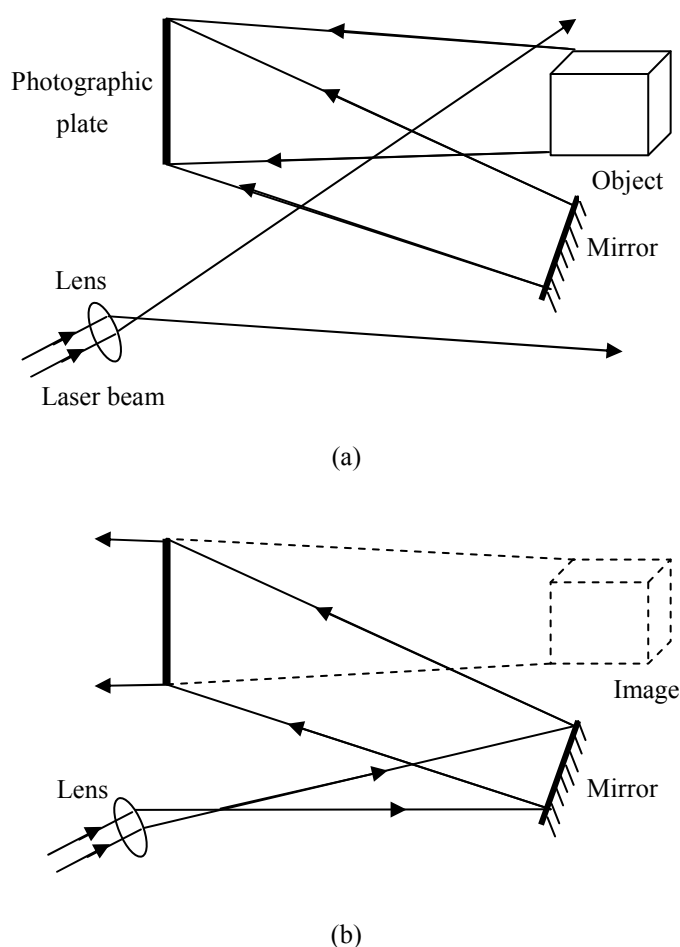


Fig. 1.2 The basic procedures of holography: (a) hologram recording; (b) hologram reconstruction

As Fig. 1.2 shows, holography consists of two procedures: recording and reconstruction. In Fig. 1.2(a), a hologram is recorded on a high resolution photographic plate. What is actually recorded is the interference pattern formed by the light scattered from the surface of the object and a coherent beam which is called the reference wave. Here, the requirement of a

coherent beam ensures that the interference pattern is displayed stably in time. The object cannot be recognized directly from its hologram, which is encoded in the form of bright and dark microscopic fringes. These micro fringes usually are not visible to the human eye, due to the high spatial frequencies involved. The hologram is usually recorded on a flat surface, but it contains 3D information. When a sufficiently long exposure time has elapsed, depending upon the laser power employed and the sensitivity of the film/plate, the photographic film is developed by a wet chemical method.

Subsequently the hologram is placed back in the recording geometry as shown in Fig. 1.2(b) and it is illuminated with a reference wave similar to the original reference wave that was used to record the hologram. A virtual image is then located in the same position that was previously occupied by the object. This virtual image, formed by part of the transmitted light, exhibits all the normal attributes of perspective and depth of focus that the object would exhibit if it were still there.

To illustrate the principle of holography more clearly, the holographic process may be described mathematically. Assume that the coordinate system of the hologram plane is the xy plane. In the procedure of recording the hologram, the complex amplitude of the object wave can be written as

$$U_o(x, y) = E_o(x, y) \exp[i\phi_o(x, y)] \quad (1.1)$$

where E_o is the real amplitude and ϕ_o is the phase of the object wave. Similarly, the reference wave can be written as

$$U_r(x, y) = E_r(x, y) \exp[i\phi_r(x, y)] \quad (1.2)$$

where E_r is the real amplitude and ϕ_r is the phase of the reference wave. When both waves interfere with each other at the surface of the recording medium, the intensity is given by:

$$\begin{aligned} U_h(x, y) &= |U_o(x, y) + U_r(x, y)|^2 \\ &= [U_o(x, y) + U_r(x, y)] \cdot [U_o(x, y) + U_r(x, y)]^* \\ &= |U_o(x, y)|^2 + |U_r(x, y)|^2 + U_o(x, y)U_r^*(x, y) + U_o^*(x, y)U_r(x, y) \end{aligned} \quad (1.3)$$

where $*$ means complex conjugate. The amplitude transmission $I(x, y)$ of the developed photographic plate is proportional to $U_h(x, y)$:

$$I(x, y) = h_0 + \beta\tau U_h(x, y) \quad (1.4)$$

The constant β is dependent on the exposure characteristics of the light sensitive material, τ is the exposure time and h_0 is the amplitude transmission of the unexposed plate. $I(x, y)$ is called the hologram function. From Eq. (1.1), Eq. (1.2), Eq. (1.3), and Eq. (1.4), the hologram function can be written as:

$$I(x, y) = h_0 + \beta\tau[E_o^2 + E_r^2 + U_o(x, y)U_r^*(x, y) + U_o^*(x, y)U_r(x, y)] \quad (1.5)$$

In the reconstruction process, the hologram function is multiplied by the complex amplitude of the original reference wave:

$$\begin{aligned} U_r(x, y)I(x, y) = & [h_0 + \beta\tau(E_o^2 + E_r^2)]U_r(x, y) \\ & + \beta\tau E_r^2 U_o(x, y) + \beta\tau U_r^2(x, y)U_o^*(x, y) \end{aligned} \quad (1.6)$$

The first term on the right hand side of Eq. (1.6) is the undiffracted wave passing through the hologram. As the hologram can be regarded as a diffraction grating, this term indicates the zero order diffraction term of the hologram. The second term is the reconstructed object wave, whose brightness is affected by the real constant $\beta\tau E_r^2$. This is a virtual image. The third term is a distorted real image of the object. For in-line holography, where the axes of the object wave and the reference wave are parallel, these three terms superpose together and cannot be separated spatially. However in off-axis holography, these three terms are spatially separated and may be isolated.

1.3 The development of optical holography

The history of holography shows a rapid development of the technology. In the limited space of this review chapter, only the major milestones are mentioned here. A more detailed introduction can be found in the bibliography (Smith, 1969, Hariharan, 1984).

The first idea of a two-stage imaging process was proposed by Bragg (1939, 1942) in his X-ray microscope. In his experiment, an X-ray diffraction pattern of the crystal lattice was first recorded photographically. Then the photograph was illuminated by a filtered mercury arc lamp to implement the second diffraction process. The core of Bragg's method is a double-diffraction process, which is also the key to the holographic process. It is noted that the diffraction field of an object can be represented as a Fourier transform of the light distribution at the object and so the first recording is a Fourier transform of the object and the

second diffraction is a Fourier transform of the Fourier transform of the object, which is an image of the object itself. If all the diffracted information of amplitudes and phase are preserved, the diffraction from the hologram will reproduce the object wave. The preservation of object phase was realized by the choice of crystals that possess a center of symmetry as they are able to produce a diffracted field that would be real. Another constriction of this method was the strong magnification differential for the different wavelengths, X-ray and visible light, used in each stage respectively.

Dennis Gabor (1948) extended this idea and built the foundation of optical holography when he tried to improve the resolution in electron microscopy. He named this new technology “holography”. This was a major innovation in the optical imaging field and later he was awarded the Nobel Prize for this invention.

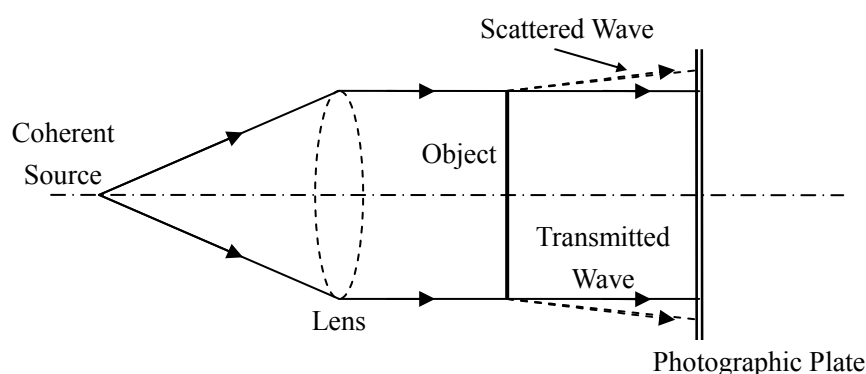


Fig. 1.3 The diagram of recording an in-line (Gabor) hologram

In the process of what is now called “in-line interferometric holography”, the photographic plate was illuminated by a plane wave through a low contrast transparent phase object, where the axes of both the object wave and the reference wave were collinear. The reference wave was used to interfere with the diffracted object wave, in order to keep the phase information of the object wave in an intensity-based interference pattern. Both waves were assumed to be mutually coherent. This interference pattern was recorded by photographic film in order to generate a hologram and the light diffracted from this hologram during reconstruction constituted the second diffraction stage, having the effect of restoring the original object wave. When Gabor applied this method to electron microscopy, it failed for various reasons. The most serious problem was the twin image in the reconstruction. As the photographic plate does not record the absolute phase, two possible objects contribute to the

single exposure that generates the hologram. One object is the original object and the other is a virtual object located symmetrically behind the hologram. Through reconstruction, the diffracted light forms two waves: the original object wave and a “twin” wave caused by the virtual object. The reconstructed object wave is so seriously disturbed by the twin wave that various researchers worked on resolving this problem (Gabor, 1949, Bragg and Rogers, 1951, Gabor, 1951, Baez, 1952). Meanwhile, the strict requirement for electrical and mechanical stability and the lack of an intense, coherent source also prevented holography from practically reaching the theoretical resolution.

Gabor’s approach and Bragg’s X-ray microscope are similar, however the differences between them are extremely significant as well. In Bragg’s microscope, no phase information is lost and an exact reconstructed image can be obtained, but the object must be symmetrical. In Gabor’s approach, an object which does not have symmetrical structure can be handled. But the loss of phase information, though tolerable, leads to the formation of an additional conjugate image.

Although some researchers, such as Rogers (1950, 1952) and Kirkpatrick and El-Sum (1956) made efforts to significantly extend the theory and conceptual understanding of holography, it was not until the early 1960s that significant further progress was made in this field, when Leith and Upatnieks (1962, 1963) proposed a new method to completely remove the twin conjugate wave. The method is simple and utilises an off-axis reference wave. The experimental geometry is shown in Fig. 1.2. From an optical viewpoint, the interference between the diffracted object wave and an off-axis reference wave would form a hologram with the structure of a diffraction grating. When light is diffracted from this hologram in the reconstruction process, it will yield two waves representing the first two orders of the grating. In this way, physical separation is achieved and the previously troublesome superposition of the twin wave is completely removed. By a simple tilting of the reference wave, dramatic improvement is achieved and this started a new era of holographic applications. Another advantage of this method is that the film can be processed with a reduced requirement for a linear transfer of exposure to amplitude transmission. It also has the ability to eliminate the self-interference effects between different points of the object. Finally, it removes the restrictions against the use of objects which do not transmit a large proportion of light.

Even though a number of advantages were provided by this new method, the real potential of research in the holographic field was not realised until another two important advances

emerged. In the early 1960s, the advent of the gas laser, which is a highly coherent source, with strong irradiance, was one of the important advances. It significantly increased the resolution of holography and gives the ability to make holograms of very large objects. Another advance by Leith and Upatnieks (1964) was diffuse illumination holography. With diffuse illumination on the object, this technology has the ability to record holograms of diffusely reflecting, 3D objects as compared to the previous technology, which only records holograms of very thin transparent objects. Diffuse illumination holography also has the capability to superpose multiple holograms on the same holographic plate. This phenomenon is very helpful enabling multi-exposure methods in the application of holographic interferometry.

Denisyuk (1962) invented the thick reflection hologram for which the images are viewed in light that is reflected from the hologram. In this case, which is shown in Fig. 1.4, the beam reflected from the object and the reference beam must arrive at the photographic plate from opposite sides to record a reflection hologram. This type of hologram has a very good wavelength selectivity which allows it to be illuminated by white light to reconstruct the images.

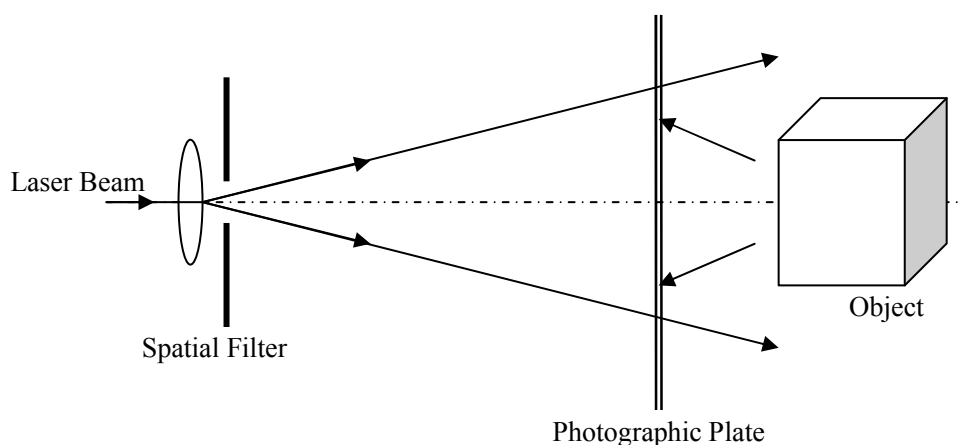


Fig. 1.4 A simple geometry of recording one-beam reflection hologram

Because of the great potential of holography, it has been applied to many scientific and engineering fields. These included high-resolution imaging of aerosols (Thompson et al., 1967), imaging through diffusing and aberrating media (Leith and Upatnieks, 1966, Kogelnik, 1965), multiple imaging (Lu, 1968, Groh, 1968), computer generated holograms (Lohmann and Paris, 1967), the production and correction of optical elements (Upatnieks et al., 1966), information storage and processing (Stroke et al., 1965, Vander Lugt et al., 1965),

etc. Among these, holographic interferometry (HI) is a major application of holography, which was discovered almost simultaneously by several groups, all working independently (Brooks et al., 1965, Burch, 1965, Collier et al., 1965, Haines and Hildebrand, 1965, Powell and Stetson, 1965). It became possible, for the first time, to map the displacements of a relatively rough surface with an accuracy of a fraction of a micrometre (Amadesi et al., 1974); it was also possible to make interferometric comparisons of stored wavefronts that existed at different times.

The explosive growth of optical holography since the advent of the laser in the 1960's was restricted with the beginning of the digital era in the late 1990s. Conventional optical holography uses photographic film as the recording medium which cannot be processed in real-time. The wet chemical procedure also adds complexity to the applications of holography. From 1994, a new type of holography, digital holography emerged and is becoming an increasingly active field for optical researchers.

1.4 The development of digital holography

Before Leith and Upatnieks proposed off-axis holography, Rogers reasoned that the interference pattern of an on-axis hologram could be calculated analytically for some simple objects, for example a wire. He fabricated the holograms and reconstructed the images. But this method could go no further until the computing power was dramatically increased and the fast Fourier transform (FFT) algorithm was invented. Brown and Lohmann (1966) were successful in using detour phase¹ and the FFT algorithm to create binary computer-generated holograms. Afterwards the computer-generated holograms were transferred to a transparency by a plotting or printing device and they were then optically reconstructed. This offers a great advantage in that the hologram of an object can be created accurately; even it never existed in the physical world. Limitations of the method include the feasibility of creating the object mathematically in the first place, the acceptable duration for the calculations involved, and the performance of the device used to transfer the numerical hologram to a suitable transparent medium. This technology is not discussed in this thesis. For a more detailed description and discussion, please read the papers in the bibliography (Lee, 1978, Bryngdahl and Wyrowski, 1990).

¹ While the path difference for wavelets from adjacent slits of a perfect grating in the first diffraction order is exactly one wavelength, the path length difference for wavelets from a dislocated slit and its neighbour will be greater or less than one wavelength. The deviation from an integral wavelength is called "detour phase", which forms the basis for encoding the phase in computer-generated holograms.

The development of computer technology made possible not only the digital recording process in the computer but also the numerical reconstruction process as well. The idea of numerical reconstruction was proposed by Goodman and Lawrence (1967) and by Kronrod et al. (1972). A hologram on a photographic plate is optically enlarged and sampled and then this digitized hologram is reconstructed numerically. Other researchers improved the reconstruction algorithm and applied it to various fields, for example particle measurement (Onural and Scott, 1987, Liu and Scott, 1987, Onural and Ozgen, 1992), shape measurement (Ma et al., 2004) and microscopy (Haddad et al., 1992).

In 1993, Schnars and Juptner (1993) used a CCD camera to record a hologram and performed numerical reconstruction in order to reconstruct this digital hologram. This represents a significant step in bringing new life to holography. It was named “direct holography” originally, but later the term “digital holography” was widely accepted by the optical metrology community for this method. It is a method that moves directly from optical recording to numerical processing. There is no need for a holographic plate or other chemical media so no wet chemical processing involved, which offers the potential of real-time applications. So far it has been successfully applied to deformation analysis (Pedrini et al., 2003), shape measurement (Yamaguchi et al., 2006), microscopy (Cuche et al., 1999), particle distribution measurement (Koek et al., 2005), light-in-flight holography and other short-coherence-length applications (Nilsson and Carlsson, 1998), comparative digital holography (Osten et al., 2002), holographic optical tweezers (Reicherter et al., 2006) and encryption of information (Meng et al., 2005), etc. A detailed introduction and discussion for the definition and applications of this method can be found in the bibliography (Schnars and Juptner, 2005).

A clear path of development has been shown above for holography, from its birth to the situation today. With the introduction of developments in electronics technology, holography has been given a new lease of life since the end of the last century. Digital holography or electronic holography adds some exciting new features to traditional optical science. It allows the simulation of experiments in a computer and even the ability to perform scenarios that cannot be realised by real experimentation. For the first time, phase information can be obtained directly, which is a significant contribution to optics.

1.5 Some problems in digital holography

Although digital holography has some significant advantages over conventional optical holography, there are some important restrictions on this new technique.

1.5.1 The spatial frequency requirements

In digital holography, a CCD/CMOS camera is used as the recording medium to take the place of the high-resolution photographic film or plates used in optical holography. The interference pattern generated by the superposition of the reference wave and the wave from the object points has to be resolved in order to restore the intensity and phase information of the object. The maximum spatial frequency that can be resolved is determined by the maximum angle θ_{\max} between these waves:

$$f_{\max} = \frac{2}{\lambda} \sin \frac{\theta_{\max}}{2} \quad (1.7)$$

Photographic emulsions used in optical holography have resolutions of up to 5000 line pairs per millimetre (Lp/mm). Using these materials, holograms with angles of up to 180° between the reference wave and the object wave can be recorded. To obtain good reconstructed image quality and measurement accuracy, each micro fringe in the hologram must be represented by at least two pixels. Then the corresponding maximum resolvable spatial frequency is calculated by:

$$f_{\max} = \frac{1}{2\Delta x} \quad (1.8)$$

However, the distance between neighbouring pixels of a high resolution CCD camera is only in the order of $\Delta x \approx 5\mu m$. According to Eq. (1.8), the maximum resolvable spatial frequency of a CCD camera is about $100 Lp/mm$. Combining Eq. (1.7) and (1.8) leads to

$$\theta_{\max} = 2 \arcsin \left(\frac{\lambda}{4\Delta x} \right) \approx \frac{\lambda}{2\Delta x} \quad (1.9)$$

where the approximation is valid for small angles. Therefore the distance between neighbouring pixels is the quantity, which limits the maximum angle between reference and object wave. For the camera which has been used in the experiments described in the latter chapters (ProgRes[®] MF^{scan}, manufactured by JENOPTIK AG), its pixel size is $6.45\mu m$ and

the maximum angle between the reference wave and the object wave should be less than 2.81° . For off-axis holography, half of the full resolution of the hologram is occupied by the twin image, which means the valid resolution for the off-axis hologram is only half of the result in Eq. (1.9).

The limited resolution of the CCD device restricts digital holography being applied to large objects or small objects that are very close to the CCD device. In digital holography, both the size of the object and the distance of the object to the CCD have to be carefully chosen to ensure the validity of the experimental results. A lot of effort has been made to increase the resolution of digital holography. In the latter part of this thesis, these issues will be discussed in detail.

1.5.2 The algorithm for numerical reconstruction

Fresnel-Kirchoff integral is used to describe the diffraction of a light wave at an aperture which is mounted perpendicular to the incoming beam. In the case of holography, the hologram is the aperture. The numerical calculation of the Fresnel-Kirchoff integral is of vital importance in performing the numerical reconstruction of the digital hologram. There are some algorithms available to realise this process, but different algorithms vary in the quality of the reconstructed image.

1.5.3 The suppression of the zero-order term and twin image

The CCD/CMOS only records the intensity of the interference pattern, as the photographic film does in optical holography. Similarly to what was illustrated in section 1.3, after numerical reconstruction in the computer, it will be the case that the zero-order term and the twin image also exist together with the desired image. Due to the limited resolution and dynamic range of CCD sensors, the zero-order term and the twin image will severely affect the quality of the reconstructed object image. They must be suppressed in order to gain useful results from digital holography.

1.6 Objective of this thesis

The objective of this thesis is to explore the feasibility of digital holography and apply it to the specific application of optical contouring of macroscopic objects.

1.7 Contributions of this thesis

- Investigated the algorithms for performing numerical reconstruction, especially the Fresnel approximation and convolution approach, and successfully applied them to both simulated ideal holograms and real holograms obtained from the optical system.
- Comprehensively investigated the methods to suppress the zero-order term and the twin image; made the comparisons with different methods to find out the suitable solution for the objective of this thesis; pointed out the benefits and limitations of each method; proposed a digital filtering method to suppress the zero-order term and the twin image of the holograms of the diffusely reflective object.
- Explored the methods to enhance the reconstruction results of the digital holograms, including optimising the intensity ratio of the object illumination wave and the reference wave, the reduction of speckle noise and the improvement of the resolution of the reconstruction results; presented a method to get the suitable intensity ratio between the object wave and the reference wave; successfully applied a super resolution (SR) image reconstruction technique to digital holography in order to suppress the aliasing effect caused by the insufficient resolution of the CCD device.
- Performed optical contouring of 3D objects via the two-source technique and two-wavelength technique; analysed every aspect of performing two-source contouring technique based on digital holography in detail; applied SR technique to two-source contouring, and achieved larger measurement range than that without using SR technique; proved the feasibility of the application of SR technique in phase measurement of digital holographic interferometry.
- Proposed the use of two-source contouring of an object of accurately known dimensions as a numerical metric to evaluate the effects of different processing techniques used in digital holography.

1.8 The structure of this thesis

In Chapter 1 of this thesis, the background of the techniques for shape measurement is given. Among the various techniques, digital holography is a new branch that is full of potential in this field. A historical review of holography has shown the roots of holography itself and the subsequent development of digital holography. The problems to be investigated in this thesis have been proposed. Also the objectives and contributions of this work have been presented.

In Chapter 2, the different algorithms for the numerical reconstruction of digital holograms are discussed. Results based on these algorithms are presented and analysed. Comparisons of the different algorithms are also presented.

In Chapter 3, available techniques to suppress the zero-order term and twin image are described. The results of simulations and experiments are shown in this chapter. Analyses of these methods and comparisons between the methods are presented in detail.

In Chapter 4, the factors which affect the quality of the reconstructed images are discussed and optimized, including approaches to overcoming the restriction of the relatively low CCD sampling rate. A new type of camera is used to generate ultra high resolution images, i.e. super images, to increase the resolution of the digital hologram and these “super resolution” methods are analysed in detail.

In Chapter 5, digital holography has applied to the shape measurement of macroscopic objects. The problems involved with real measurements are revealed and solutions are proposed. The experimental results for some real objects are shown.

In Chapter 6, the conclusions from this thesis are presented. Some suggestions on future work in the area of digital holography are presented.

References:

- ABRAMSON, N. (1976) Holographic contouring by translation. *Applied Optics*, 15, 1018-22.
- AMADESI, S., GORI, F., GRELLA, R. & GUATTARI, G. (1974) Holographic methods for painting diagnostics. *Applied Optics*, 13, 2009-2013.
- BAEZ, A. V. (1952) Study in diffraction microscopy with special reference to X-rays. *Journal Of The Optical Society Of America*, 42, 756-762.
- BRAGG, W. L. (1939) A new type of 'X-ray microscope'. *Nature*, 143, 678.
- BRAGG, W. L. (1942) The X-ray microscope. *Nature*, 149, 470-471.
- BRAGG, W. L. & ROGERS, G. L. (1951) Elimination of the unwanted image in diffraction microscopy. *Nature*, 167, 190-191.
- BROOKS, R. E., HEFLINGER, L. O. & WUERKER, R. F. (1965) Interferometry with a holographically reconstructed comparison beam. *Applied Physics Letters*, 7, 248-249.
- BROWN, B. R. & LOHMANN, A. W. (1966) Complex spatial filtering with binary masks. *Applied Optics*, 5, 967-969.
- BURCH, J. M. (1965) The application of lasers in production engineering. *The Production Engineer*, 44, 431-442.
- CHEN, F., BROWN, G. M. & MUMIN, S. (2000) Overview of three-dimensional shape measurement using optical methods. *Optical Engineering*, 39, 10-22.
- COLLIER, R. J., DOHERTY, E. T. & PENNINGTON, K. S. (1965) Application of moire techniques to holography. *Applied Physics Letters*, 7, 223-225.
- CUCHE, E., MARQUET, P. & DEPEURSINGE, C. (1999) Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms. *Applied Optics*, 38, 6994-7001.
- DENISYUK, Y. N. (1962) Photographic reconstruction of the optical properties of an object in its own scattered radiation field. *Soviet Physics - Doklady*, 7, 543-545.
- DRESEL, T., HAUSLER, G. & VENZKE, H. (1992) Three-dimensional sensing of rough surfaces by coherence radar. *Applied Optics*, 31, 919-25.
- GABOR, D. (1948) A new microscopic principle. *Nature*, 161, 777-118.

- GABOR, D. (1949) Microscopy by reconstructed wavefronts. *Proceedings of the Royal Society A*, 197, 454-487.
- GABOR, D. (1951) Microscopy by reconstructed wavefronts. II. *Proceedings of the Royal Society (Lond.) B*, 64, 449-469.
- GOODMAN, J. W. & LAWRENCE, R. W. (1967) Digital image formation from electronically detected holograms. *Applied Physics Letters*, 11, 77-79.
- GROH, G. (1968) Multiple imaging by means of point holograms. *Applied Optics*, 7, 1643-1644.
- HADDAD, W. S., CULLEN, D., SOLEM, J. C., LONGWORTH, J. W., MCPHERSON, A., BOYER, K. & RHODES, C. K. (1992) Fourier-transform holographic microscope. *Applied Optics*, 31, 4973-8.
- HAINES, K. & HILDEBRAND, B. P. (1965) Contour generation by wavefront reconstruction. *Physics Letters*, 19, 10 - 11.
- JI, Z. & LEU, M. C. (1989) Design of optical triangulation devices. *Opt. Laser Technol.*, 21, 335-338.
- KEFERSTEIN, C. P. & MARXER, M. (1998) Testing bench for laser triangulation sensors. *Sens. Rev.*, 18, 183-187.
- KIRKPATRICK, P. & EL-SUM, H. M. A. (1956) Image formation by reconstructed wavefronts. I. physical principles and methods of refinement. *Journal Of The Optical Society Of America*, 46, 825-831.
- KOEK, W. D., BHATTACHARYA, N., BRAAT, J. J. M., OOMS, T. A. & WESTERWEEL, J. (2005) Influence of virtual images on the signal-to-noise ratio in digital in-line particle holography. *Optics Express*, 13, 2578-2589.
- KOGELNIK, H. (1965) Holographic image projection through inhomogeneous media. *Bell System Technical Journal*, 44, 2451-2455.
- KRONROD, M. A., MERZLYAKOV, N. S. & YAROSLAVSKY, L. P. (1972) Reconstruction of holograms with a computer. *Soviet Physics-Technical Physics*, 17, 333-334.
- KUJAWINSKA, M., SALBUT, L. & PATORSKI, K. (1991) Three-channel phase stepped system for moire interferometry. *Applied Optics*, 30, 1633-5.

- LEITH, E. N. & UPATNIEKS, J. (1962) Reconstructed wavefronts and communication theory. *Journal of the Optical Society of America*, 52, 1123-1130.
- LEITH, E. N. & UPATNIEKS, J. (1963) Wavefront reconstruction with continuous-tone objects. *Journal Of The Optical Society Of America*, 53, 1377-1381.
- LEITH, E. N. & UPATNIEKS, J. (1964) Wavefront reconstruction with diffused illumination and three-dimensional objects. *Journal Of The Optical Society Of America*, 54, 1295-1301.
- LEITH, E. N. & UPATNIEKS, J. (1966) Holographic imagery through diffusing media. *Journal Of The Optical Society Of America*, 56, 523.
- LIU, G. & SCOTT, P. D. (1987) Phase retrieval and twin-image elimination for in-line Fresnel holograms. *Journal Of The Optical Society Of America A*, 4, 159-165.
- LOHMANN, A. W. & PARIS, D. P. (1967) Binary Fraunhofer holograms generated by computer. *Applied Optics*, 6, 1739-1748.
- LU, S. (1968) Generating multiple images for integrated circuits by Fourier-transform holograms. *Proceedings of the IEEE*.
- MA, L., WANG, H., LI, Y. & JIN, H. (2004) Numerical reconstruction of digital holograms for three-dimensional shape measurement. *Journal of Optics A: Pure and Applied Optics*, 6, 396-400.
- MENG, X. F., CAI, L. Z., HE, M. Z., DONG, G. Y. & SHEN, X. X. (2005) Cross-talk-free double-image encryption and watermarking with amplitude-phase separate modulations. *Journal of Optics A: Pure and Applied Optics*, 7, 624-631.
- NILSSON, B. & CARLSSON, T. E. (1998) Direct three-dimensional shape measurement by digital light-in-flight holography. *Applied Optics*, 37, 7954-9.
- ONURAL, L. & OZGEN, M. T. (1992) Extraction of three-dimensional object-location information directly from in-line holograms using Wigner analysis. *Journal Of The Optical Society Of America A*, 9, 252-260.
- ONURAL, L. & SCOTT, P. D. (1987) Digital decoding of in-line holograms. *Optical Engineering*, 26, 1124-1132.
- OSTEN, W., BAUMBACH, T. & JUPTNER, W. (2002) Comparative digital holography. *Optics Letters*, 27, 1764-1766.

- PEDRINI, G., GUSEV, M., SCHEDIN, S. & TIZIANI, H. J. (2003) Pulsed digital holographic interferometry by using a flexible fiber endoscope. *Optics and Lasers in Engineering*, 40, 487-99.
- POWELL, R. L. & STETSON, K. A. (1965) Interferometric vibration analysis by wavefront reconstruction. *Journal Of The Optical Society Of America*, 55, 1593-1598.
- REICHERTER, M., ZWICK, S., HAIST, T., KOHLER, C., TIZIANI, H. & OSTEN, W. (2006) Fast digital hologram generation and adaptive force measurement in liquid-crystal-display-based holographic tweezers. *Applied Optics*, 45, 888-896.
- ROGERS, G. L. (1950) Gabor diffraction microscopy: the hologram as a generalized zone-plate. *Nature*, 166, 237.
- ROGERS, G. L. (1952) Experiments in diffraction microscopy. *Proceedings of the Royal Society (Edinburgh)*, A63, 193-221.
- SCHNARS, U. & JUPTNER, W. (1993) Principles of direct holography for interferometry. IN JUPTNER, W. & OSTEN, W. (Eds.) *FRINGE 93 Proceeding of 2nd International Workshop on Automatic Processing of Fringe Patterns*. Akademie, Berlin.
- SRINIVASAN, V., LIU, H. C. & HALIOUA, M. (1984) Automated phase-measuring profilometry of 3-D diffuse objects. *Applied Optics*, 23, 3105-8.
- STROKE, G. W., RESTRICK, R., FUNKHOUSER, A. & BRUMM, D. (1965) Resolution-retrieving compensation of source effects by correlation reconstruction in high resolution holography. *Physics Letters*, 18, 274-275.
- THOMPSON, B. J., WARD, J. H. & ZINKY, W. R. (1967) Application of hologram techniques for particle size analysis. *Applied Optics*, 6, 519-526.
- TIZIANI, H. J. (1997) Optical metrology of engineering surfaces--scope and trends. IN RASTOGI, P. K. (Ed.) *Optical Measurement Techniques and Applications*. Boston, Artech House.
- UPATNIEKS, J., VANDER LUGT, A. & LEITH, E. N. (1966) Correction of lens aberrations by means of holograms. *Applied Optics*, 5, 589-593.
- VAN HAASTEREN, A. J. P. & FRANKENA, H. J. (1994) Real-time displacement measurement using a multicamera phase-stepping speckle interferometer. *Applied Optics*, 33, 4137-42.

- VANDER LUGT, A., ROTZ, F. B. & KLOOSTER JR, A. (1965) Character reading by optical spatial filtering. IN TIPPETT, J. T., BERKOWITZ, D. A., CLAPP, L. C., KOESTER, C. J. & VANDERBURGH JR, A. (Eds.) *Optical & electro-optical information proceeding*. Cambridge: Massachusetts Institute of Technology Press.
- WILSON, T. (1990) *Confocal Microscopy*, London, Academic Press.
- YAMAGUCHI, I., IDA, T., YOKOTA, M. & YAMASHITA, K. (2006) Surface shape measurement by phase-shifting digital holography with a wavelength shift. *Applied Optics*, 45, 7610-7616.
- ZELENKA, J. S. & VARNER, J. R. (1969) Multiple-index holography contouring. *Applied Optics*, 8, 1431-4.

2 Numerical Reconstruction in Digital Holography

2.1 General principle of digital holography

Digital holography became feasible with the emergence of high resolution CCD/CMOS devices and the increasing availability of affordable high speed computing. The basic principle of digital holography is the same as in conventional holography, which was discussed in Chapter 1. It also consists of two procedures: a recording stage and a reconstruction stage. However in digital holography, the process of recording the hologram uses a CCD/CMOS camera, replacing photographic film as the recording medium. The parameter h_0 in Eq. (1.4) can be neglected for a CCD/CMOS camera. The CCD/CMOS camera transfers the captured image to the computer connected to and this image is saved digitally as a digital hologram. In the process of reconstruction, this hologram is digitally accessed and numerically reconstructed by a virtual reference wave which effectively simulates the reference wave used in the process of recording. As the reconstruction procedure is performed in the computer, its speed only depends upon the implementation of the numerical reconstruction algorithm and the speed of the computer processing.

Because the reference wave has to be generated virtually in the computer, a plane wave or a spherical wave is usually used in the recording process, for the sake of simplicity and accuracy. The object is either transparent or a 3D body with a diffusely reflecting surface, which is at a certain distance from the CCD/CMOS device. Fig. 2.1 shows the typical set-up of digital holography.

A hologram can be looked upon as an aperture, and the reconstructed images are the results of diffraction of the reference wave by this aperture. Huygen's principle illustrates that every point of a wavefront can be considered as a source point for secondary spherical waves. The wavefront at any other place is the coherent superposition of these secondary waves. Suppose the coordinate system is as shown as in Fig. 2.2(a), then the diffraction by the aperture or hologram in the distance of d along the propagation direction of the wave can be quantitatively described by Fresnel-Kirchhoff integral (Schnars and Juptner, 2005):

$$\Gamma(\xi', \eta') = \frac{i}{\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x, y) \frac{\exp\left(-i \frac{2\pi}{\lambda} \rho'\right)}{\rho'} Q dx dy \quad (2.1)$$

with

$$\rho' = \sqrt{(x - \xi')^2 + (y - \eta')^2 + d^2} \quad (2.2)$$

and

$$Q = \frac{(\cos \theta + \cos \theta')}{2} \quad (2.3)$$

In Eq. (2.1), $A(x, y)$ is the complex amplitude in the plane of the aperture. λ is the wavelength used in the diffraction. $\Gamma(\xi', \eta')$ is the diffracted field in the observation plane. ρ' represents the Euclidean distance between a point in the aperture plane and a point in the observation plane which can be calculated by Eq. (2.2). Q stands for the inclination factor which depends on the angles θ and θ' according to Eq. (2.3). θ is the angle between the incident ray from the source and the unit vector \vec{n} perpendicular to the aperture plane, and θ' is the angle between the diffracted ray and \vec{n} , see Fig. 2.2(b).

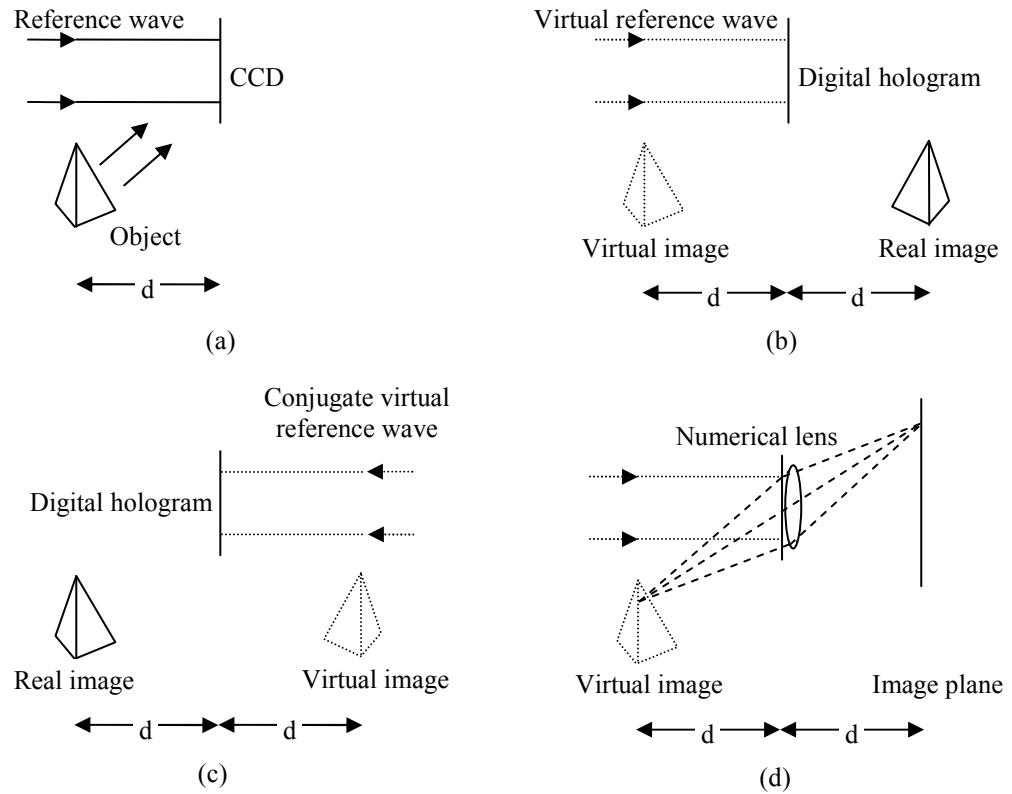


Fig. 2.1 Principle of digital holography: (a) recording; (b) reconstruction with reference wave; (c) reconstruction with conjugate reference wave; (d) reconstruction of the virtual image

Some authors (see for example Goodman, 2005) use a “+” sign in the argument of the exponential function in Eq. (2.1) instead of the “-” sign used here. Which is appropriate depends on the mathematical description of the optical wave which can be defined as either $\exp(+i\varphi)$ or $\exp(-i\varphi)$. However, using the “+” sign in Eq. (2.1) leads to the same expressions for all measurable quantities, as e.g. the intensity and the magnitude of the interference phase used in digital holographic interferometry.

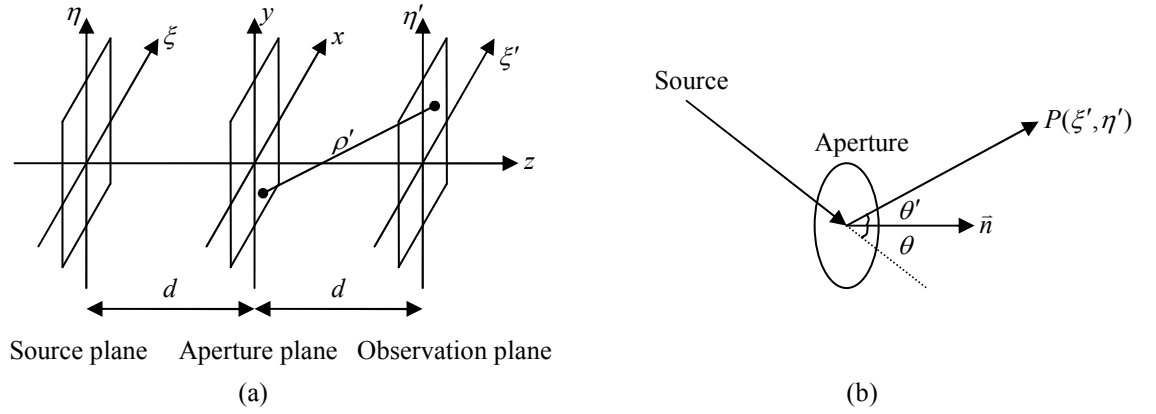


Fig. 2.2 Illustrations for diffraction: (a) coordinate system; (b) angles in inclination factor Q

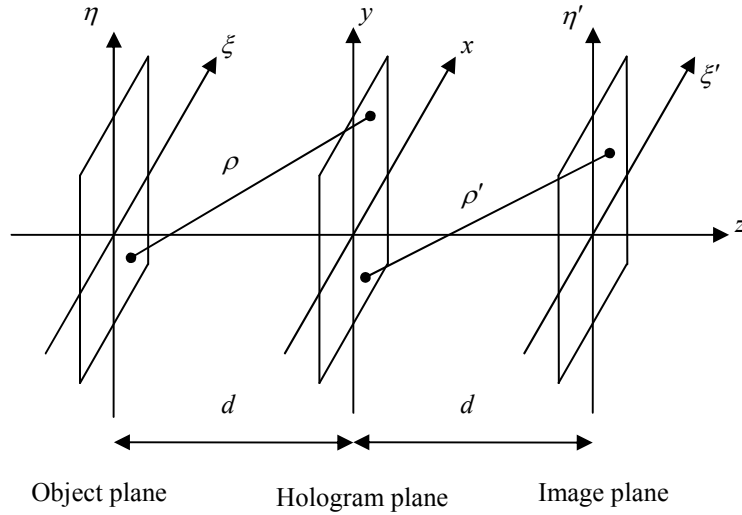


Fig. 2.3 Coordinate system for numerical hologram reconstruction

The formulation above covers the conventions in the case of classical diffraction within an aperture. In the case of digital holography, the coordinate system is quite similar to Fig. 2.2(a) but with a different meaning. For the sake of clarity, the coordinate system for numerical reconstruction is shown in Fig. 2.3. The inclination factor Q is set to 1 because the angles θ and θ' are approximately zero in practical digital holography – a

consequence of the relatively low resolution of the detector. This is valid for all numerical reconstruction algorithms in this thesis.

If the reference wave is set up to be nominally normally incident to the hologram, then the diffracted light is approximated by the Fresnel-Kirchhoff integral as

$$\Gamma(\xi', \eta') = \frac{i}{\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_h(x, y) U_r(x, y) \frac{\exp\left(-i \frac{2\pi}{\lambda} \rho'\right)}{\rho'} dx dy \quad (2.4)$$

where $U_h(x, y)$ is the digital hologram captured by the CCD camera; λ is the wavelength of the light in the virtual reference beam used in the reconstruction; ρ' is the distance between a point in the hologram plane and a point in the reconstruction plane which has the same form as Eq. (2.2); $U_r(x, y)$ is the function describing the reference wave. For a uniform plane reference wave, it has real amplitude with the simple equation:

$$U_r(x, y) = a_r + i0 = a_r \quad (2.5)$$

For a spherical reference wave whose point source is at the original point of the object plane, it can be described as:

$$U_r(x, y) = \frac{\exp\left(-i \frac{2\pi}{\lambda} \sqrt{d^2 + x^2 + y^2}\right)}{\sqrt{d^2 + x^2 + y^2}} \approx \frac{1}{d} \exp\left(-i \frac{2\pi}{\lambda} d\right) \exp\left[-i \frac{\pi}{\lambda d} (x^2 + y^2)\right] \quad (2.6)$$

In Eq. (2.4), $\Gamma(\xi', \eta')$ is the diffraction pattern calculated at a distance d behind the CCD plane (see Fig. 2.3) which means it reconstructs the complex amplitude in the plane of the real image. Therefore, both the intensity and the phase information can be obtained after numerical reconstruction. However, in optical reconstruction, only the intensity is visible. So this new facility potentially sheds new light in many applications of holography. The reconstructed intensity is written as:

$$I(\xi', \eta') = |\Gamma(\xi', \eta')|^2 \quad (2.7)$$

And the reconstructed phase is:

$$\varphi(\xi', \eta') = \arctan \frac{\text{Im}[\Gamma(\xi', \eta')]}{\text{Re}[\Gamma(\xi', \eta')]} \quad (2.8)$$

where Re denotes the real part and Im denotes the imaginary part. Please notice here the calculated diffraction pattern is the complex amplitude at a distance d behind the CCD plane where the real image is reconstructed.

However as was mentioned in section 1.2 the real image could be distorted by the reference wave (see Eq. (1.6)). To avoid this effect and ensure that an undistorted real image are left, a conjugate reference wave has to be introduced in the reconstruction as shown in Fig. 2.1(c). Contrary to the situation depicted in Fig. 2.1(b), an undistorted real image swaps with the virtual image and is located at the position where the object was recorded. Then the calculated diffraction pattern is rewritten as:

$$\Gamma(\xi, \eta) = \frac{i}{\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_h(x, y) U_r^*(x, y) \frac{\exp\left(-i \frac{2\pi}{\lambda} \rho\right)}{\rho} dx dy \quad (2.9)$$

with

$$\rho = \sqrt{(x - \xi)^2 + (y - \eta)^2 + d^2} \quad (2.10)$$

where $U_r^*(x, y)$ is conjugate to the original reference wave $U_r(x, y)$. But for the plane reference wave defined in Eq. (2.5), both results from Eq. (2.4) and Eq. (2.9) are equivalent because $U_r(x, y) = U_r^*(x, y)$.

The virtual image can be reconstructed by adding a numerical lens in the process of numerical reconstruction. The simplest scheme is shown in Fig. 2.1(d) where the numerical lens is located directly behind the hologram. The introduced lens functions as an eyepiece for the observer, viewing the optical reconstruction. Suppose the lens has a focal length of f , the imaging formula of the lens is:

$$L(x, y) = \exp\left[i \frac{\pi}{\lambda f} (x^2 + y^2)\right] \quad (2.11)$$

To keep the magnification to 1, $f = d/2$ is usually used in Eq. (2.11). The phase aberrations caused by the introduced lens should be corrected to avoid phase errors in the reconstructed image. The phase correction factor is described as:

$$P(\xi', \eta') = \exp\left[i \frac{\pi}{\lambda f} (\xi'^2 + \eta'^2)\right] \quad (2.12)$$

According to Eq. (2.4), Eq. (2.11), and Eq. (2.12), the full equation for virtual image reconstruction through a numerical lens of focal length f is:

$$\Gamma(\xi', \eta') = \frac{i}{\lambda} P(\xi', \eta') \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_h(x, y) U_r(x, y) L(x, y) \frac{\exp\left(-i \frac{2\pi}{\lambda} \rho'\right)}{\rho'} dx dy \quad (2.13)$$

This combination of conventional holography and electronic devices allows holograms to be taken and reconstructed in near real time. With the use of CCD/CMOS cameras, the hologram can be recorded directly and stored digitally. No wet-chemical or other time-consuming processes are required. Numerical reconstruction also offers greater flexibility than optical processing. Without using the phase shifting methods, the phases of the stored light waves can be calculated directly from the digital holograms under numerical reconstruction, as shown in Eq. (2.8). Also other optical metrology methods, for example, shearography, can be derived numerically from digital holography (Schnars and Juptner, 1994). In the next section, the numerical methods used to calculate the diffracted field, which is the key element in digital holography, will be discussed in detail.

2.2 Numerical reconstruction of digital holograms

As illustrated above, Eq. (2.9) is the key formula of digital holography and it is essential to calculate it numerically to perform numerical reconstruction of a digital hologram. The direct approach of Eq. (2.9) is not feasible in terms of the calculation complexity and computer run time. Some approximations have to be applied in order to calculate the quadrature of this double integral to make the numerical reconstruction effective and efficient. According to the approximation used in the algorithm, the numerical reconstruction can be classified into three types: Fresnel approximation, convolution approaches, and other numerical methods.

2.2.1 Reconstruction by the Fresnel approximation

In digital holography, the values of the coordinates x and y as well as ξ and η , are very small compared to the distance d between the reconstruction plane and the CCD/CMOS device. If now the right hand side of Eq. (2.10) is expanded to a Taylor series and the fourth term is smaller than the wavelength,

$$\frac{[(x - \xi)^2 + (y - \eta)^2]^2}{8d^3} \ll \lambda \quad (2.14)$$

the effect of it and the terms after it are negligible and they can be removed. Thus the distance ρ can be approximated as:

$$\rho = d + \frac{(\xi - x)^2}{2d} + \frac{(\eta - y)^2}{2d} \quad (2.15)$$

Replacing the denominator in Eq. (2.9) with d and inserting Eq. (2.15) into it, the following expression results in the reconstruction of the real image:

$$\begin{aligned} \Gamma(\xi, \eta) = & \frac{i}{\lambda d} \exp\left(-i \frac{2\pi}{\lambda} d\right) \exp\left[-i \frac{\pi}{\lambda d} (\xi^2 + \eta^2)\right] \\ & \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_r^*(x, y) U_h(x, y) \exp\left[-i \frac{\pi}{\lambda d} (x^2 + y^2)\right] \exp\left[i \frac{2\pi}{\lambda d} (x\xi + y\eta)\right] dx dy \end{aligned} \quad (2.16)$$

This equation is known as the Fresnel approximation or Fresnel transformation due to its mathematical similarity with the Fourier transform.

Similarly to Eq. (2.7) and Eq. (2.8), the intensity is calculated by squaring:

$$I(\xi, \eta) = |\Gamma(\xi, \eta)|^2 \quad (2.17)$$

And the phase is calculated by

$$\varphi(\xi, \eta) = \arctan \frac{\text{Im}[\Gamma(\xi, \eta)]}{\text{Re}[\Gamma(\xi, \eta)]} \quad (2.18)$$

To convert the Fresnel approximation in Eq. (2.16) to a digital implementation, two substitutions are applied (Yarovslavsky and Merzlyakov, 1980):

$$\nu = \frac{\xi}{\lambda d}; \quad \mu = \frac{\eta}{\lambda d} \quad (2.19)$$

Therefore Eq. (2.16) turns into:

$$\begin{aligned}
\Gamma(\nu, \mu) &= \frac{i}{\lambda d} \exp\left(-i \frac{2\pi}{\lambda} d\right) \exp\left[-i \pi \lambda d (\nu^2 + \mu^2)\right] \\
&\quad \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_r^*(x, y) U_h(x, y) \exp\left[-i \frac{\pi}{\lambda d} (x^2 + y^2)\right] \exp[i 2\pi (x \nu + y \mu)] dx dy \\
&= \frac{i}{\lambda d} \exp\left(-i \frac{2\pi}{\lambda} d\right) \exp\left[-i \pi \lambda d (\nu^2 + \mu^2)\right] \\
&\quad \times \mathfrak{F}^{-1}\left\{U_r^*(x, y) U_h(x, y) \exp\left[-i \frac{\pi}{\lambda d} (x^2 + y^2)\right]\right\}
\end{aligned} \tag{2.20}$$

This shows that the Fresnel approximation is equal to the multiplication of a spherical phase factor and the inverse Fourier transformation of the function $U_r^*(x, y) U_h(x, y) \exp[-i \pi (x^2 + y^2)/(\lambda d)]$.

Suppose that there is a rectangular raster of $N \times N$ points with steps Δx and Δy along the coordinate axes. Δx and Δy are the spacings of two adjacent pixels on the CCD/CMOS device in the horizontal and vertical directions, respectively. The calculated diffraction function $\Gamma(\nu, \mu)$ is sampled with this raster so that the infinite integral in Eq. (2.20) is converted to finite sums, as given by:

$$\begin{aligned}
\Gamma(m, n) &= \frac{i}{\lambda d} \exp\left(-i \frac{2\pi}{\lambda} d\right) \exp\left[-i \pi \lambda d (m^2 \Delta \nu^2 + n^2 \Delta \mu^2)\right] \\
&\quad \times \sum_k^{N-1} \sum_l^{N-1} U_r^*(k, l) U_h(k, l) \exp\left[-i \frac{\pi}{\lambda d} (k^2 \Delta x^2 + l^2 \Delta y^2)\right] \exp[i 2\pi (k \Delta x m \Delta \nu + l \Delta y n \Delta \mu)]
\end{aligned} \tag{2.21}$$

for $m = 0, 1, \dots, N-1$; $n = 0, 1, \dots, N-1$.

Because the maximum frequency is determined by the sampling interval in the spatial domain according to the theory of the Fourier transform,

$$\nu_{\max} = N \Delta \nu = 1/\Delta x; \quad \mu_{\max} = N \Delta \mu = 1/\Delta y \tag{2.22}$$

The relationships among Δx , Δy , $\Delta \nu$ and $\Delta \mu$ are:

$$\Delta \nu = \frac{1}{N \Delta x}; \quad \Delta \mu = \frac{1}{N \Delta y} \tag{2.23}$$

With Eq. (2.23), Eq. (2.21) can be rewritten as:

$$\begin{aligned}
\Gamma(m, n) &= \frac{i}{\lambda d} \exp\left(-i \frac{2\pi}{\lambda} d\right) \exp\left[-i \pi \lambda d \left(\frac{m^2}{N^2 \Delta x^2} + \frac{n^2}{N^2 \Delta y^2}\right)\right] \\
&\quad \times \sum_k^{N-1} \sum_l^{N-1} U_r^*(k, l) U_h(k, l) \exp\left[-i \frac{\pi}{\lambda d} (k^2 \Delta x^2 + l^2 \Delta y^2)\right] \exp\left[i 2\pi \frac{(km + ln)}{N}\right] \\
&= \frac{i}{\lambda d} \exp\left(-i \frac{2\pi}{\lambda} d\right) \exp\left[-i \pi \lambda d \left(\frac{m^2}{N^2 \Delta x^2} + \frac{n^2}{N^2 \Delta y^2}\right)\right] \\
&\quad \times \mathfrak{T}^{-1} \left\{ U_r^*(k, l) U_h(k, l) \exp\left[-i \frac{\pi}{\lambda d} (k^2 \Delta x^2 + l^2 \Delta y^2)\right] \right\}
\end{aligned} \tag{2.24}$$

Eq. (2.24) is known as the discrete Fresnel transform. The matrix Γ is calculated by applying an inverse discrete Fourier transform to the product of $U_r^*(k, l)$ with $U_h(k, l)$ and $\exp[-i\pi(k^2 \Delta x^2 + l^2 \Delta y^2)/(\lambda d)]$. The calculation can be done very effectively using the Fast Fourier Transform (FFT) algorithm. If only the intensity according to Eq. (2.17) is of interest, the factor in front of the sum only affects the overall phase and can be neglected. If the wavelength does not change in multiple exposures, this factor can also be removed when calculating the phase differences between holograms.

From Eq. (2.19) and Eq. (2.23), the sizes of the reconstructed pixels, $\Delta\xi$ and $\Delta\eta$, are different from the sampled pixel in the hologram, Δx and Δy . The relationship between them is given as:

$$\Delta\xi = \frac{\lambda d}{N \Delta x}; \quad \Delta\eta = \frac{\lambda d}{N \Delta y} \tag{2.25}$$

Through a similar process, the corresponding discrete formula for reconstruction via a virtual lens with focal length $f = d/2$ is:

$$\begin{aligned}
\Gamma(m, n) &= \frac{i}{\lambda d} \exp\left(-i \frac{2\pi}{\lambda} d\right) \exp\left[i \pi \lambda d \left(\frac{m^2}{N^2 \Delta x^2} + \frac{n^2}{N^2 \Delta y^2}\right)\right] \\
&\quad \times \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} U_r(k, l) U_h(k, l) \exp\left[i \frac{\pi}{\lambda d} (k^2 \Delta x^2 + l^2 \Delta y^2)\right] \exp\left[i 2\pi \left(\frac{km}{N} + \frac{ln}{N}\right)\right]
\end{aligned}$$

for $m = 0, 1, \dots, N-1$; $n = 0, 1, \dots, N-1$

(2.26)

Some examples both via computer simulation and experiment can now be given to demonstrate the process of numerical reconstruction by the Fresnel approximation.

A simulation is shown in Fig. 2.4. The original object, a letter “F”, is shown in Fig. 2.4(a). Random phase is applied to this object. The hologram is generated by the interference between the diffraction field of this object on the hologram plane and a plane reference wave normally incident to the hologram plane. The wavelength in the simulation is $632.8nm$. The distance between the object plane and the hologram plane is set at $1m$. The hologram plane is an array of 512×512 pixels with a pixel spacing of $\Delta x = \Delta y = 0.01mm$. According to these parameters and Eq. (2.25), the object plane is sampled by $\Delta \xi = \Delta \eta = 0.124mm$. The plane reference wave is described as in Eq. (2.5). There are two ways to calculate the diffraction field of the object numerically.

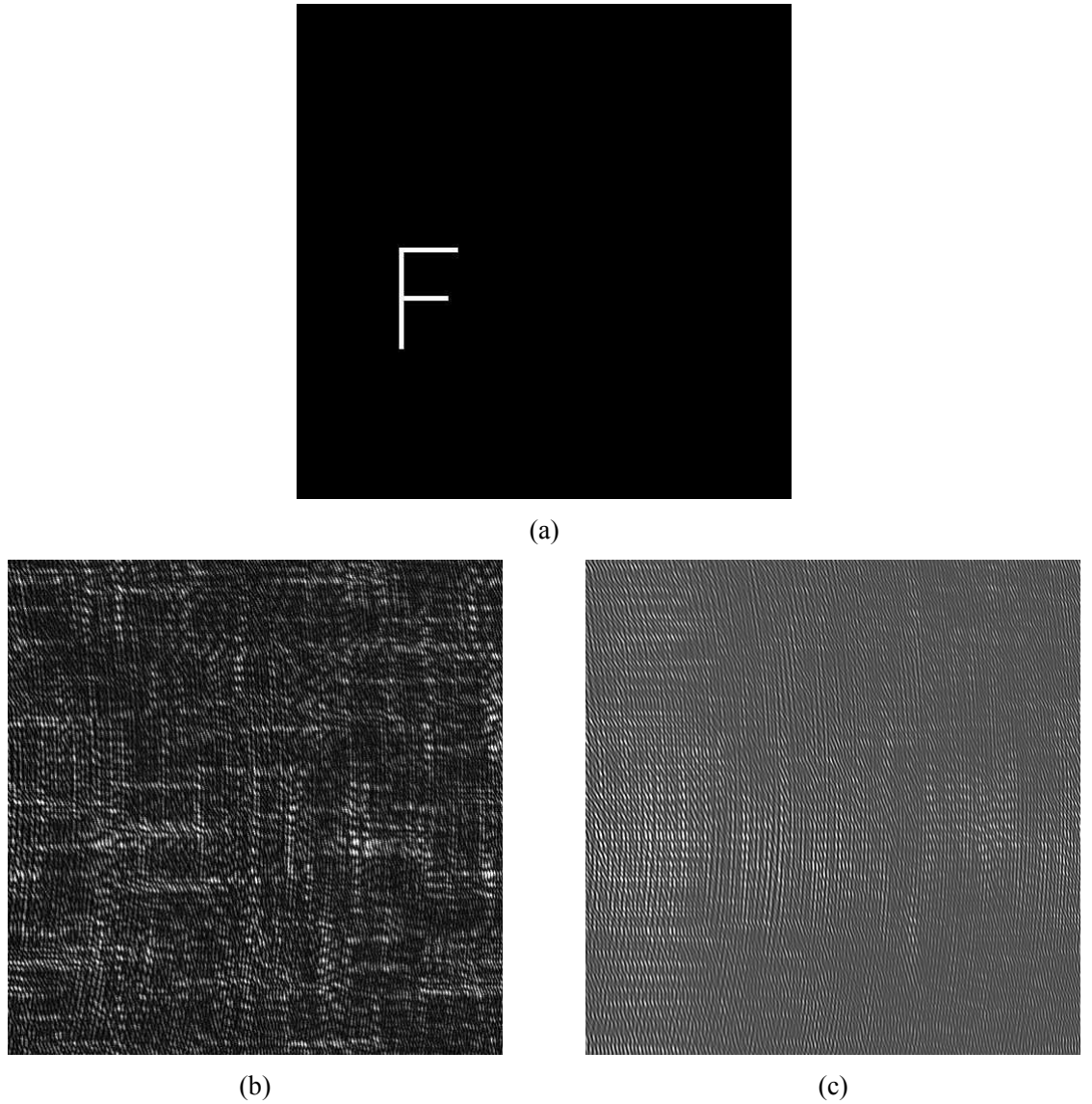


Fig. 2.4 Simulation results of numerical reconstruction by Fresnel approximation: (a) the 2D object used to generate the hologram; (b) the hologram generated by direct approach in Eq. (2.27); (c) the hologram generated by approximation in Eq. (2.29).

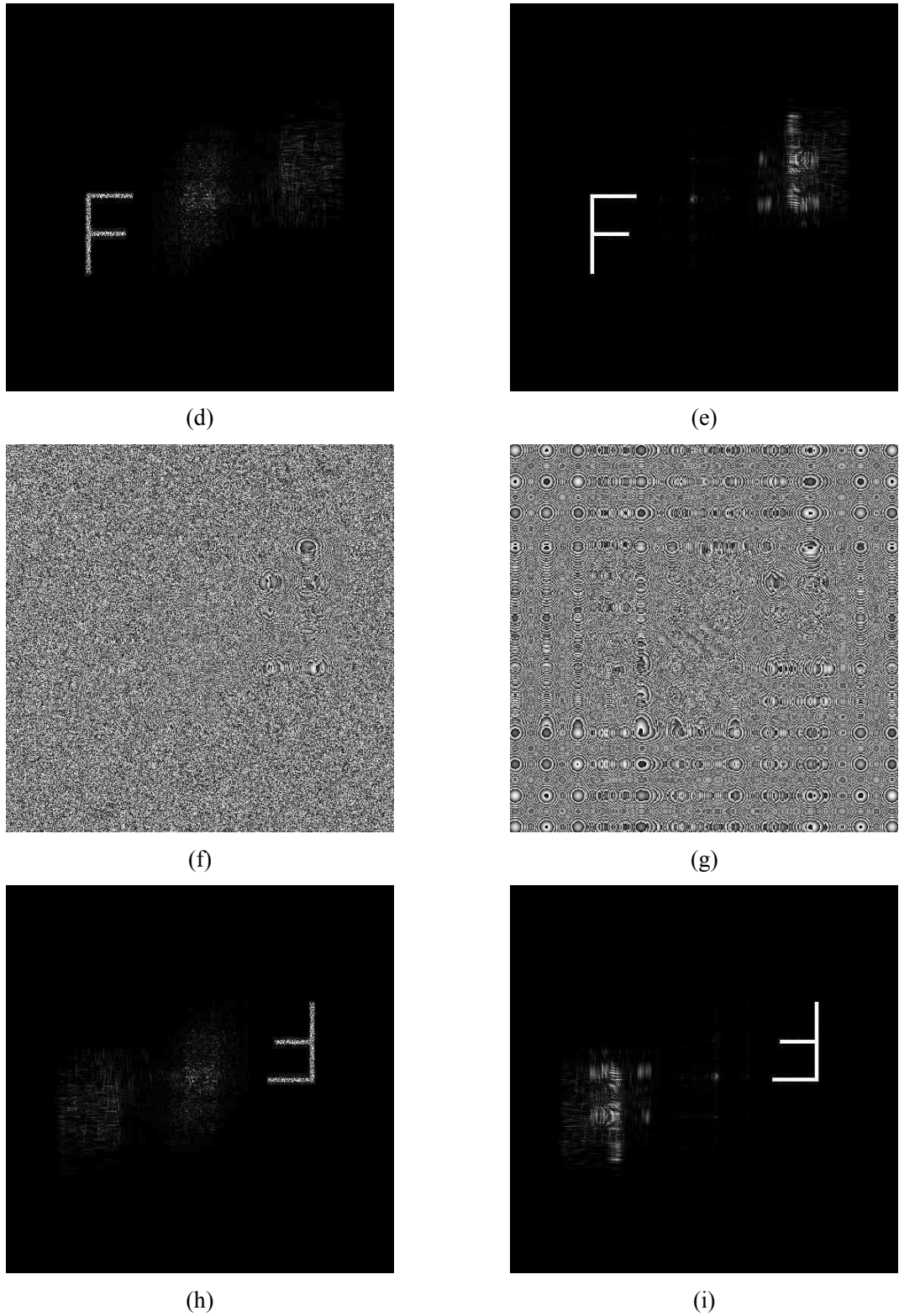


Fig. 2.4 (Continued) (d) the reconstructed real image from (b); (e) the reconstructed real image from (c); (f) the reconstructed phase image from (b); (g) the reconstructed phase image from (c); (h) the reconstructed virtual image from (b) by using a virtual lens; (i) the reconstructed virtual image from (c) by using a virtual lens.

The first method is known as the direct approach. As the optical system is a linear invariant system, the object can be decomposed to many point sources, each with different intensity

and phase, according to the definition of the object. The diffraction field of the object is the sum of the diffraction fields of those point sources in the hologram plane, according to Huygens' Principle in section 2.1. Therefore the direct approach to calculate the diffraction field of the object is derived by Eq. (2.1):

$$U_o(m\Delta x, n\Delta y) = \frac{i}{\lambda} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} U_o(k\Delta \xi, l\Delta \eta) \frac{\exp\left(-i \frac{2\pi}{\lambda} \rho\right)}{\rho}$$

for $m = 0, 1, \dots, N-1$; $n = 0, 1, \dots, N-1$.

(2.27)

$$\rho = \sqrt{(m\Delta x - k\Delta \xi)^2 + (n\Delta y - l\Delta \eta)^2 + d^2}$$

(2.28)

The coordinate system is as shown in Fig. 2.3. $U_o(k\Delta \xi, l\Delta \eta)$ is the complex amplitude of the object in the object plane; $U_o(m\Delta x, n\Delta y)$ is the diffraction field of the object in the hologram plane; λ is the wavelength in the recording process; d is the distance between the object plane and the hologram plane; Δx and Δy are the pixel spacings in the hologram plane; $\Delta \xi$ and $\Delta \eta$ are the pixel spacings in the object plane. Fig. 2.4(b) shows the computer generated hologram produced by this direct approach.

The second method is to employ an approximation of Eq. (2.28) as given by Eq. (2.15). Fig. 2.4(c) shows the computer generated hologram produced by this approximation. The diffraction field of the object then can be calculated by FFT to increase the speed of simulation dramatically. The process is written as

$$U_o(m\Delta x, n\Delta y) = \frac{i}{\lambda d} \exp\left(-i \frac{2\pi}{\lambda} d\right) \exp\left[-i \pi \lambda d (m^2 \Delta x^2 + n^2 \Delta y^2)\right] \\ \times \mathfrak{F}^{-1} \left\{ U_o(k\Delta \xi, l\Delta \eta) \exp\left[-i \frac{\pi}{\lambda d} (k^2 \Delta x^2 + l^2 \Delta y^2)\right] \right\}$$

for $m = 0, 1, \dots, N-1$; $n = 0, 1, \dots, N-1$.

(2.29)

In the reconstructed images in Fig. 2.4(d) and (e), as well as the desired image of the object, the twin image is also present on the other side of the picture but it is out of focus. For the sake of clarity, the zero-order term has already been suppressed. The methods and benefits of doing this zero-order term suppression will be presented and discussed in the next chapter. Comparing the results in Fig. 2.4, the reconstruction resulting from the hologram

in Fig. 2.4(c), using the approximation of Eq. (2.28), is better than that based on the direct method hologram of Fig. 2.4(b). However, it should be noted that Fig. 2.4(c) was obtained by applying the same approximation used in the reconstruction of the Fresnel approximation. In fact, Fig. 2.4(b) is closer to the real case and the reconstruction results bear greater resemblance to those found in experimental practice. This can be verified by considering the experimental results presented later.

Now move on to consider an actually experimental case. The optical geometry used to record a reflective digital hologram is shown in Fig. 2.5.

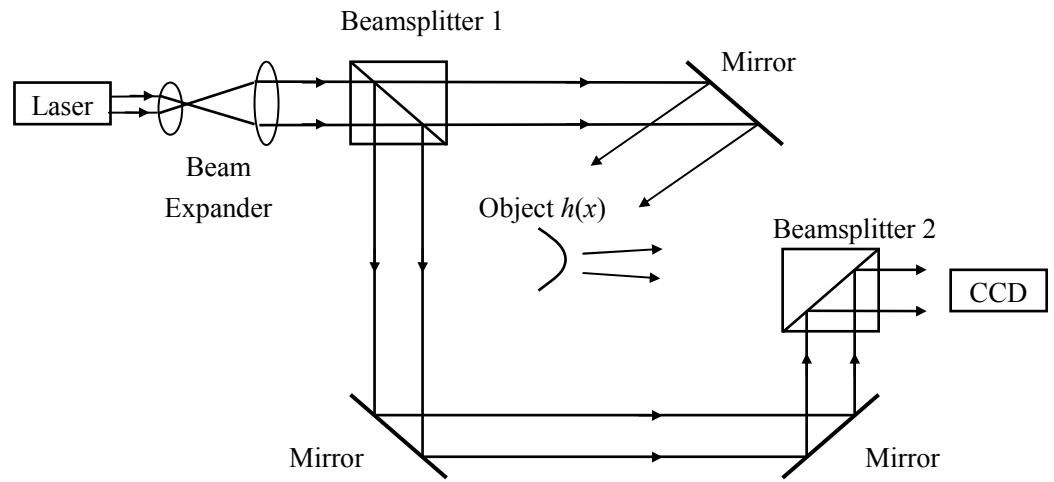


Fig. 2.5 The optical geometry to record the reflective digital hologram

A typical hologram obtained by this system is displayed in Fig. 2.6(b). The object which will be the subject of the hologram is a die with reflective surfaces and dimensions of $13mm \times 13mm \times 13mm$ as shown in Fig. 2.6(a). It is placed $580mm$ away from the CCD camera. The resolution of the camera is 1360×1024 pixels with a pixel spacing of $\Delta x = \Delta y = 6.45 \mu m$. The wavelength of the laser is $632.8nm$. Beamsplitter 1 is used to split the laser beam into two beams: a reference beam and an object beam to illuminate the die. Beamsplitter 2 recombines both beams and directs them to the CCD plane where they interfere with each other and so produced the hologram.

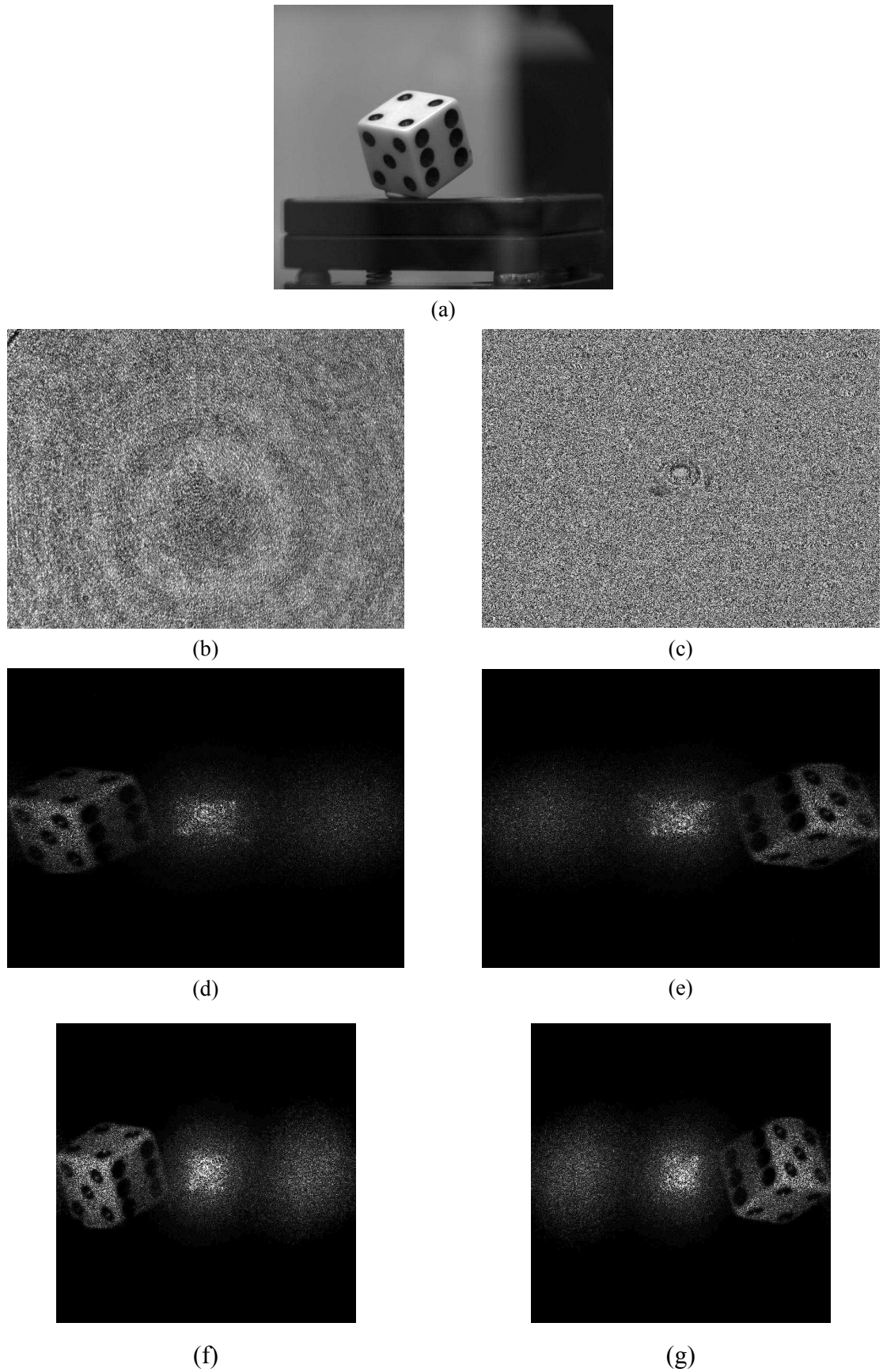


Fig. 2.6 Experimental results of numerical reconstruction by Fresnel approximation: (a) the object used to make the hologram; (b) the hologram captured by CCD; (c) the reconstructed phase image; (d) the reconstructed real image; (e) the reconstructed virtual image produced by using a virtual lens; (f) the reconstructed real image with aspect ratio of 1:1; (g) the reconstructed virtual image with aspect ratio of 1:1.

To display the real image after numerical reconstruction, the zero-order term has been suppressed. But its influence can still be seen in the reconstructed intensity images. For example in Fig. 2.6(d), the real image is displayed on the left, while the twin image on the right is out of focus, and the suppressed zero-order image is in the centre. Fig. 2.6(c) displays the reconstructed phase-contrast image of the die. It is not possible to get the phase information from this image because the surface roughness of the die is higher than the recording wavelength. But for some microscopic applications with tiny smooth objects, it is possible to simultaneously obtain the amplitude and phase information in one exposure (Cuche et al., 1999). According to Eq. (2.25), the pixel distances in the reconstructed image $\Delta\xi$ and $\Delta\eta$ are related to the pixel number and pixel distances of the CCD and the distance used in the numerical reconstruction. As the camera used in this experiment has different numbers of pixels in the ξ and η axes, the pixel distances $\Delta\xi$ and $\Delta\eta$ are different, which results in the distortion shown in Fig. 2.6(d) and (e). The image of the die does not look like the original object shown in Fig. 2.6(a). To remove this distortion from the reconstructed images, the hologram in Fig. 2.6(b) is cropped to an array of 1024×1024 thus making the pixel distances in both directions equal. The corresponding results are displayed in Fig. 2.6(f) and (g).

Fig. 2.7 shows the optical geometry required to record a digital hologram of a transparent object. The object is a USAF 1951 chromium positive resolution test target. Several groups of bars with different spacings are deposited on a glass plate as shown in Fig. 2.8(a). The target was placed 266mm away from the CCD camera. The resolution of the camera was 1360×1024 pixels, with pixel distances of $\Delta x = \Delta y = 6.45\mu\text{m}$. The wavelength of the laser was 632.8nm . Beamsplitter 1 is used to split the laser beam into two beams, the reference beam and the object beam to illuminate the resolution test target. Beamsplitter 2 recombines both beams and directs them to the CCD plane to interfere with each other and generate the hologram.

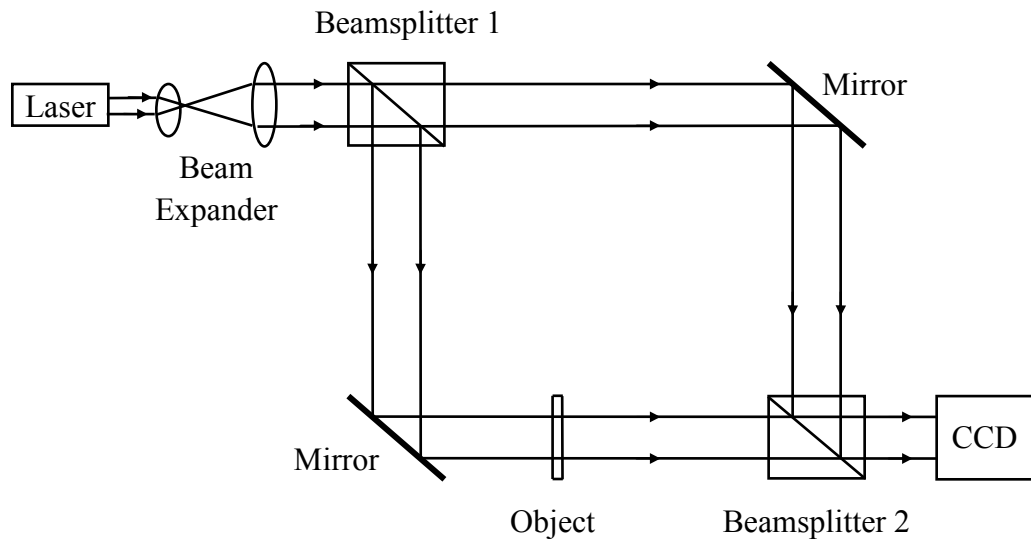


Fig. 2.7 The optical geometry to record the transparent digital hologram

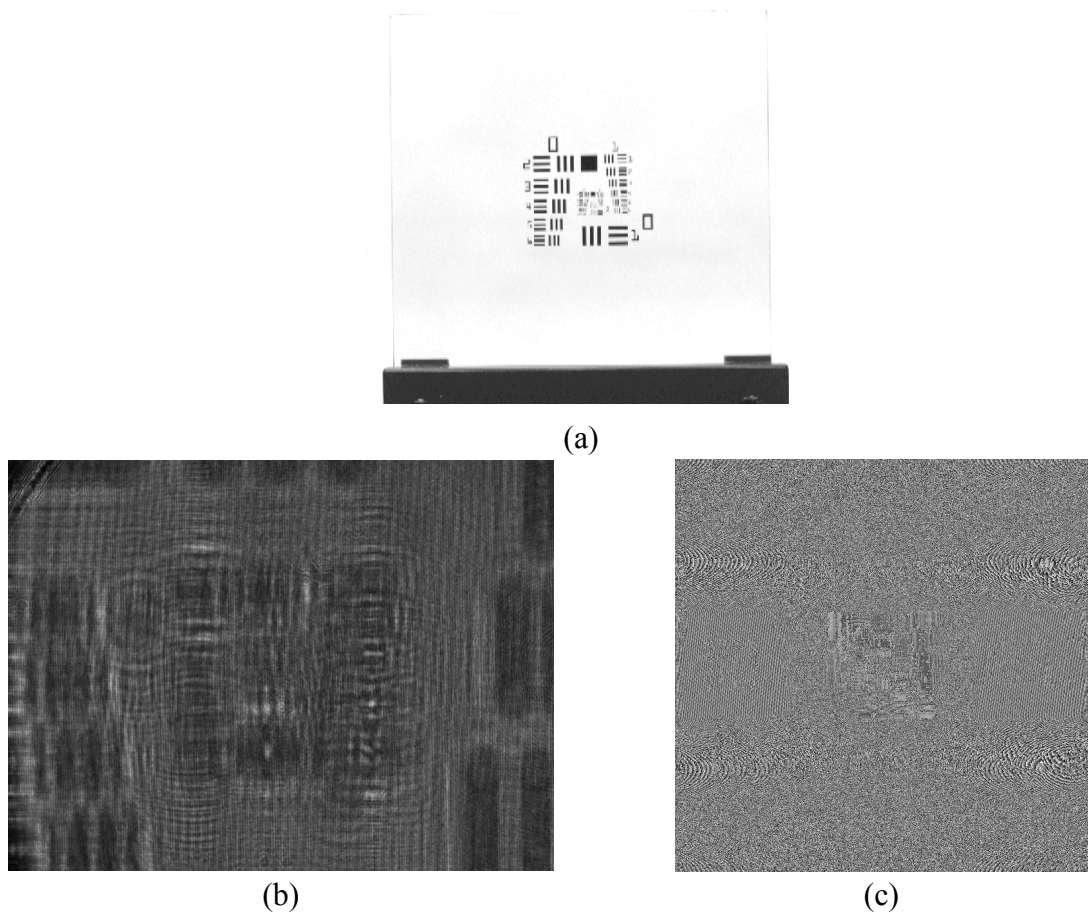


Fig. 2.8 Experimental results of numerical reconstruction by Fresnel approximation: (a) the resolution test target; (b) the hologram captured by CCD; (c) the reconstructed phase image with aspect ratio 1:1.

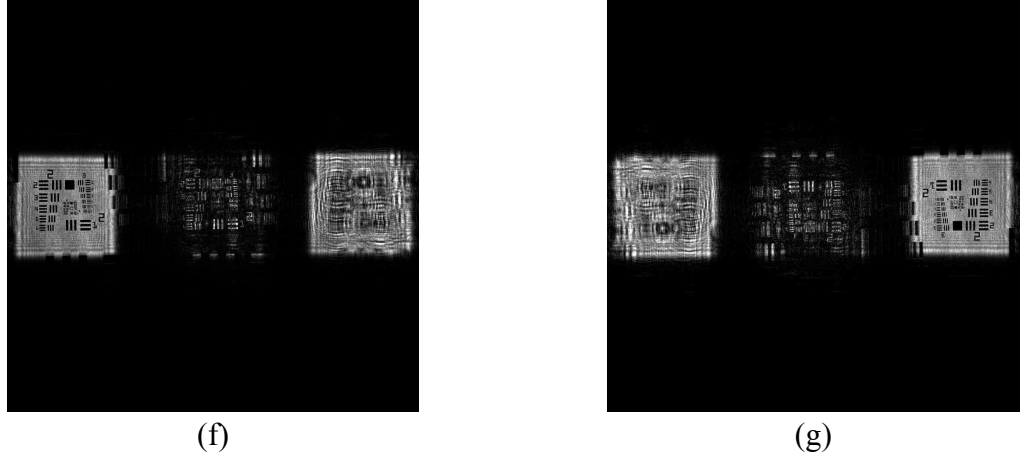


Fig. 2.8 (Continued) (d) the reconstructed real image with aspect ratio of 1:1; (e) the reconstructed virtual image with aspect ratio 1:1.

2.2.2 Reconstruction by the convolution approach

Besides the Fresnel approximation, another approach is also widely used in digital holography to perform numerical reconstruction. This method makes use of the convolution theorem. It was first applied to process the numerical reconstruction of sub-optical holograms by Demetrakopoulos and Mitta (1974). Later, this approach was introduced to digital holography by Kreis and Juptner (1997).

Eq. (2.9) also can be interpreted as a superposition integral:

$$\Gamma(\xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_h(x, y) U_r^*(x, y) g(\xi, \eta, x, y) dx dy \quad (2.30)$$

where the impulse response $g(\xi, \eta, x, y)$ is given by

$$g(\xi, \eta, x, y) = \frac{i}{\lambda} \frac{\exp \left[-i \frac{2\pi}{\lambda} \sqrt{d^2 + (x - \xi)^2 + (y - \eta)^2} \right]}{\sqrt{d^2 + (x - \xi)^2 + (y - \eta)^2}} \quad (2.31)$$

According to Eqs. (2.30) and (2.31), the linear system characterized by $g(\xi, \eta, x, y) = g(\xi - x, \eta - y)$ is space-invariant. So Eq. (2.30) can be regarded as a convolution and the convolution theorem can be applied. The two dimensional convolution of two functions $f(x, y)$ and $g(x, y)$ is defined as:

$$(f \otimes g)(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi, \eta) g(x - \xi, y - \eta) dx dy \quad (2.32)$$

where the \otimes denotes the convolution operation. The convolution theorem states that the Fourier transform of the convolution of two functions is equal to the product of the Fourier transforms of the individual functions:

$$\mathfrak{F}\{f(x, y) \otimes g(x, y)\} = \mathfrak{F}\{f(x, y)\} \mathfrak{F}\{g(x, y)\} = F(u, v) G(u, v) \quad (2.33)$$

In other words, the convolution of two functions in the spatial domain can be easily obtained through the multiplication of them in another domain, namely spatial frequency domain.

Applying the convolution theorem to Eq. (2.9), it is converted to:

$$\Gamma(\xi, \eta) = \mathfrak{F}^{-1}\left\{\mathfrak{F}\left(U_h(x, y) \cdot U_r^*(x, y)\right) \cdot \mathfrak{F}(g(x, y))\right\} \quad (2.34)$$

Eq. (2.34) includes two forward Fourier transformations and one inverse Fourier transformation, all of which can be practically implemented via the FFT algorithm. Considering the digitization, the numerical impulse response function is described as in Eq. (2.35). The shift of the coordinates by $N/2$ is to make the reconstructed area symmetrical with respect to the optical axis.

$$g(k, l) = \frac{i}{\lambda} \frac{\exp\left[-i \frac{2\pi}{\lambda} \sqrt{d^2 + \left(k - \frac{N}{2}\right)^2 \Delta x^2 + \left(l - \frac{N}{2}\right)^2 \Delta y^2}\right]}{\sqrt{d^2 + \left(k - \frac{N}{2}\right)^2 \Delta x^2 + \left(l - \frac{N}{2}\right)^2 \Delta y^2}} \quad (2.35)$$

The Fourier transform of $g(k, l)$ can be calculated and expressed analytically (Goodman, 2005) as Eq. (2.36):

$$G(n, m) = \exp\left\{-i \frac{2\pi d}{\lambda} \sqrt{1 - \frac{\lambda^2 \left(n + \frac{N^2 \Delta x^2}{2d\lambda}\right)^2}{N^2 \Delta x^2} - \frac{\lambda^2 \left(m + \frac{N^2 \Delta y^2}{2d\lambda}\right)^2}{N^2 \Delta y^2}}\right\} \quad (2.36)$$

This saves one Fourier transform for the reconstruction of the real image:

$$\Gamma(\xi, \eta) = \mathfrak{F}^{-1}\left\{\mathfrak{F}(U_h \cdot U_r^*) \cdot G\right\} \quad (2.37)$$

According to Eq. (2.13), the reconstruction of the virtual image can be considered as:

$$\Gamma(\xi', \eta') = P(\xi', \eta') \mathfrak{F}^{-1}\left\{\mathfrak{F}(U_h \cdot U_r \cdot L) \cdot G\right\} \quad (2.38)$$

Here the lens function $L(x,y)$ is defined as in Eq. (2.11) and the phase correction factor $P(\xi',\eta')$ is as defined in Eq. (2.12).

The size of the pixel reconstructed by this algorithm is the same as that of the CCD:

$$\Delta\xi = \Delta x; \quad \Delta\eta = \Delta y \quad (2.39)$$

If there is another area to reconstruct which is not symmetrical to the optical axis, two integer variables are introduced to shift the reconstructed area:

$$g(k+s_k, l+s_l) = \frac{i}{\lambda} \frac{\exp\left[-i \frac{2\pi}{\lambda} \sqrt{d^2 + \left(k - \frac{N}{2} + s_k\right)^2 \Delta x^2 + \left(l - \frac{N}{2} + s_l\right)^2 \Delta y^2}\right]}{\sqrt{d^2 + \left(k - \frac{N}{2} + s_k\right)^2 \Delta x^2 + \left(l - \frac{N}{2} + s_l\right)^2 \Delta y^2}} \quad (2.40)$$

The convolution approach provides the ability to change the image magnification in the reconstruction process. If both the focal length of the introduced numerical lens and the reconstruction distance, i.e. the distance between the hologram plane and the image plane, are changed according to the required magnification, the magnified reconstructed image can be obtained. So, in general, the reconstruction distance can set to:

$$d' = d \cdot m \quad (2.41)$$

where d is the recording distance and m represents the magnification factor. The focal length of the numerical lens can be calculated by the lens formula from geometrical optics:

$$f = \left(\frac{1}{d} + \frac{1}{d'}\right)^{-1} \quad (2.42)$$

Then the lens function in Eq. (2.11) becomes:

$$L(x,y) = \exp\left[i \frac{\pi}{\lambda} \left(\frac{1}{d} + \frac{1}{d'}\right) (x^2 + y^2)\right] \quad (2.43)$$

Applying Eq. (2.41) for d' and the lens function in Eq. (2.43) to Eq. (2.38) yields the magnified image.

In Fig. 2.9 a computer simulation of numerical reconstruction by the convolution approach is demonstrated. The wavelength in the simulation is $632.8nm$. The distance between the object plane and the hologram plane is $100mm$. The hologram plane is an array of 512×512

pixels with a pixel spacing of $\Delta x = \Delta y = 0.01mm$. According to Eq. (2.39), the object plane is sampled by the same spacing in the hologram plane as $\Delta \xi = \Delta \eta = 0.01mm$. The hologram is generated by the interference between the diffraction field of this object on the hologram plane and a plane reference wave as given by Eq. (2.27). The plane reference wave is still as described as Eq. (2.5). The object here is similar to the one in Fig. 2.4(a) but its size is about 12 times smaller than that shown in Fig. 2.4(a) because of the different pixel distances in both cases. Fig. 2.9(b) shows the computer generated hologram of this object. Fig. 2.9(c) and (e) show the reconstructed real and virtual images of Fig. 2.9(b), respectively. Fig. 2.9(d) displays the reconstructed phase image of Fig. 2.9(b). To verify the ability to change the magnification in the reconstruction process, a magnification factor is introduced to reconstruct the hologram displayed in Fig. 2.4(b) by the convolution approach. Fig. 2.9(f) shows the result for this reconstruction with a magnification factor of 12.4, which ensures the magnified image is of the same size as the original object shown in Fig. 2.4(a). However, if another magnification factor is chosen in the reconstruction process, the replica of the image will be superposed with the desired image. This is demonstrated in Fig. 2.9(g) and (h) which show the results of the reconstruction process with magnification factors of 7 and 16, respectively.

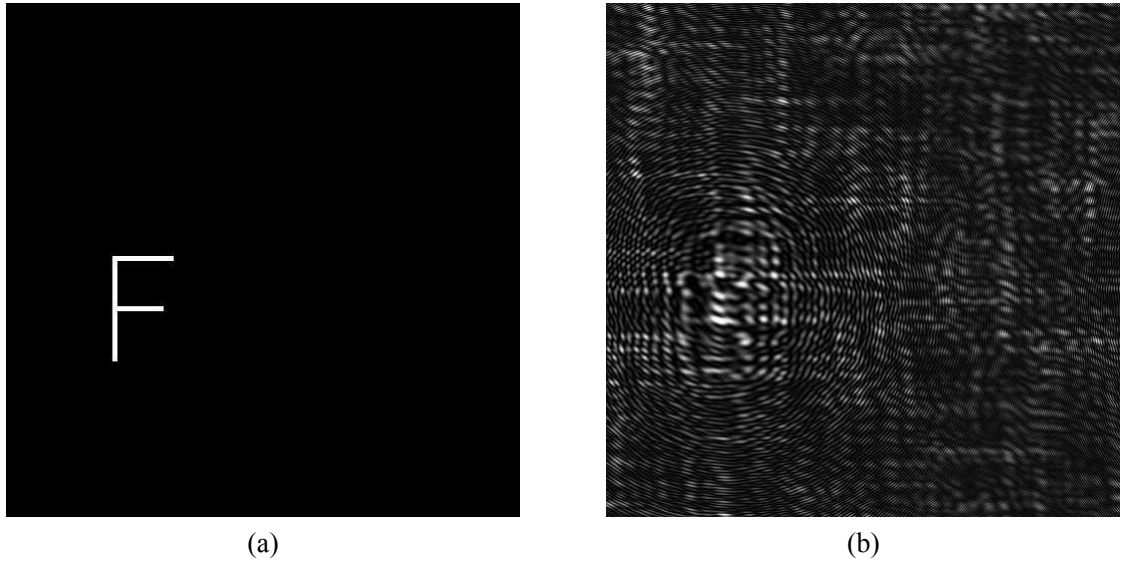


Fig. 2.9 Simulation results of the numerical reconstruction by the convolution approach: (a) the 2D object used to generate the hologram; (b) the hologram generated according to Eq. (2.27).

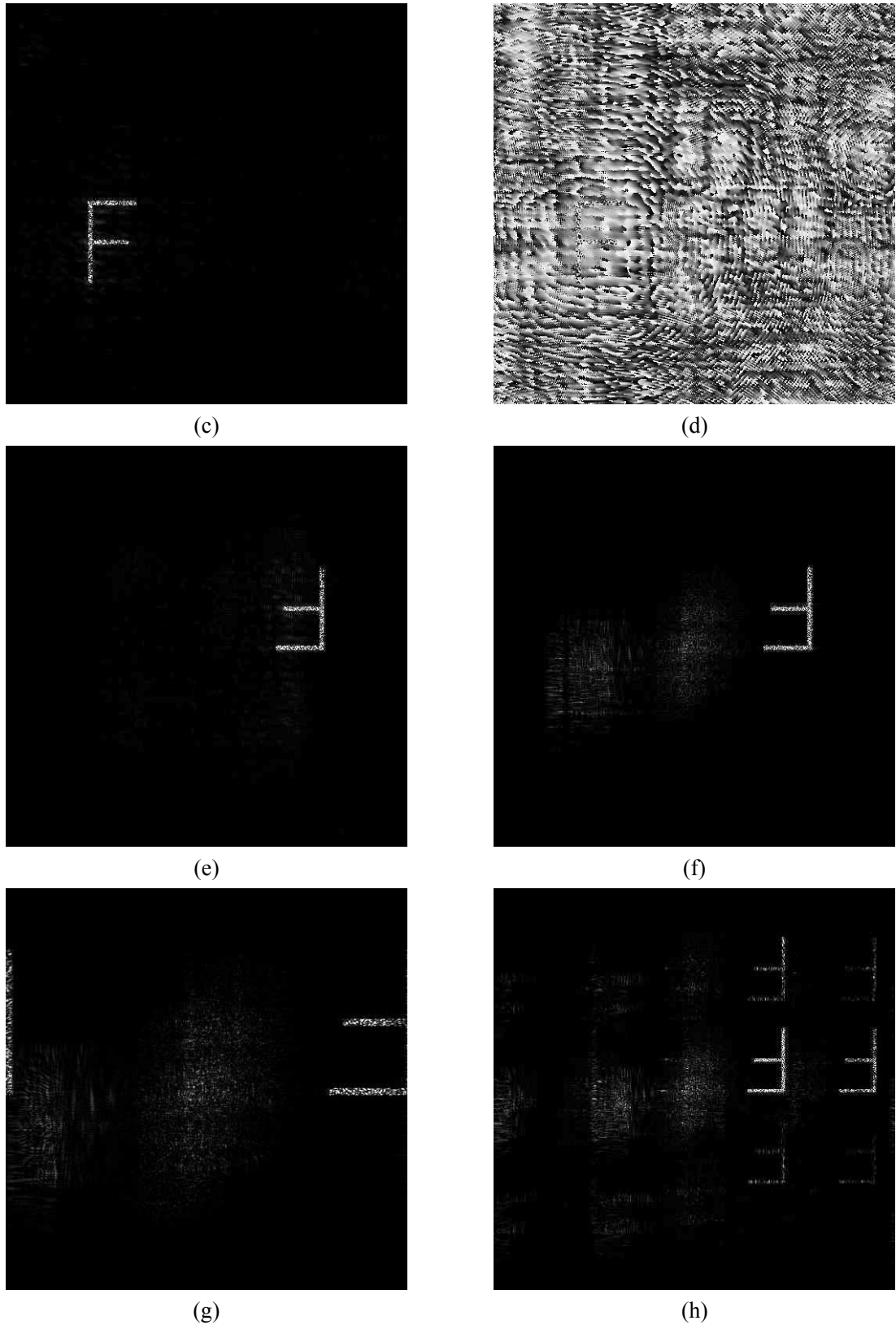


Fig. 2.9 (Continued) (c) the reconstructed real image of (b); (d) the reconstructed phase image of (b); (e) the reconstructed virtual image of (b) produced by using a virtual lens; the reconstructed virtual image of Fig. 2.4(b) produced using magnification factors of (f) 12.4, (g) 7 and (h) 16.

Experimental use of the convolution approach is demonstrated by the numerical reconstruction of the hologram shown in Fig. 2.6(b) using this method. The result is shown

in Fig. 2.10. The magnification factor was set to 6.5 which ensures that the reconstructed image is the same size as in the Fresnel approximation reconstruction. The reconstructed intensity image in Fig. 2.10(a) is comparable in quality to that obtained via previous method as shown in Fig. 2.6(f). But when the magnification factor is decreased, the whole of the object detail cannot be recovered (see in Fig. 2.10(c)); whilst when the magnification factor is increased, the unwanted replica image can be clearly seen in the reconstruction (see in Fig. 2.10(d)). These results are consistent with the computer simulation as shown previously in Fig. 2.9.

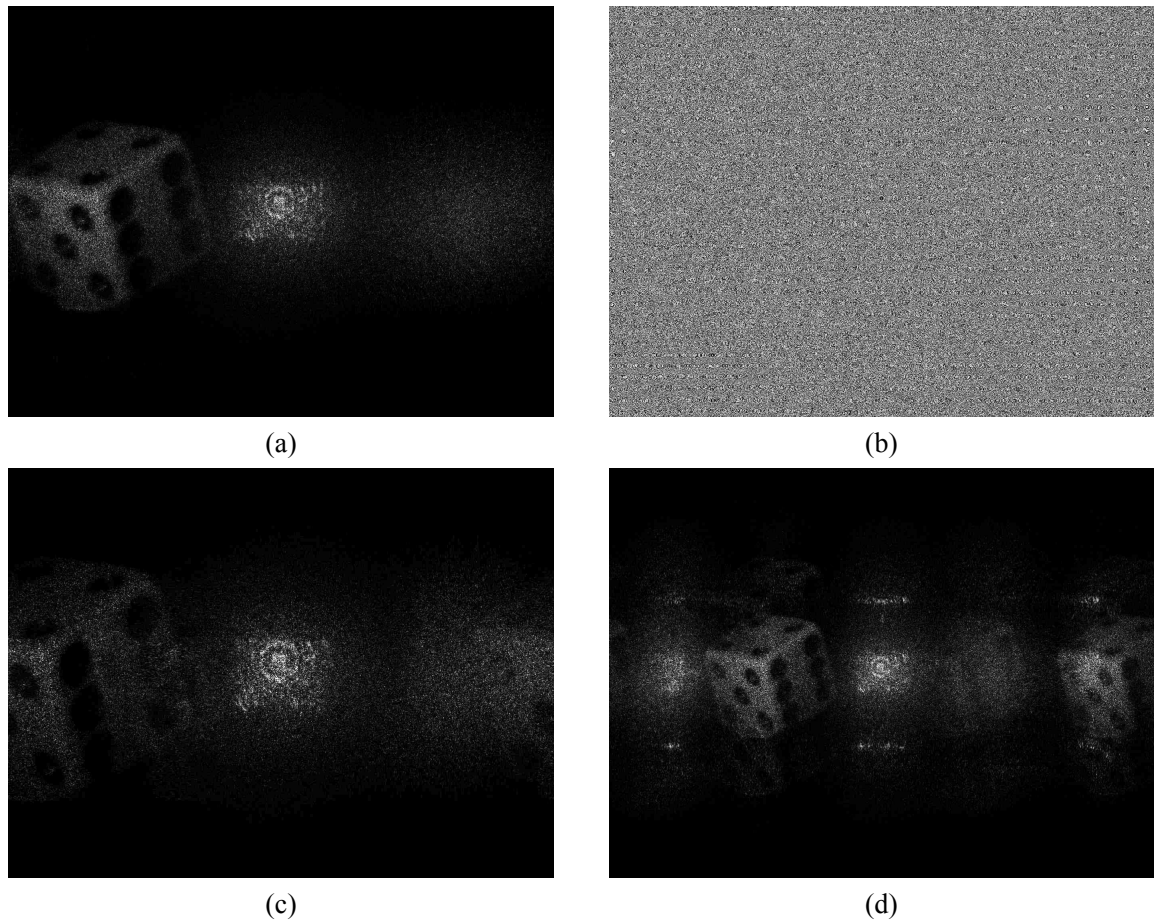
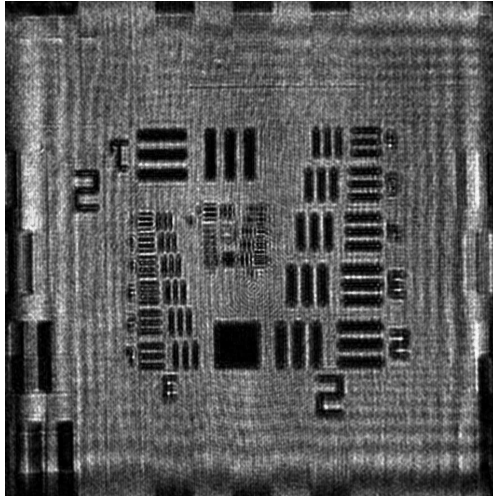


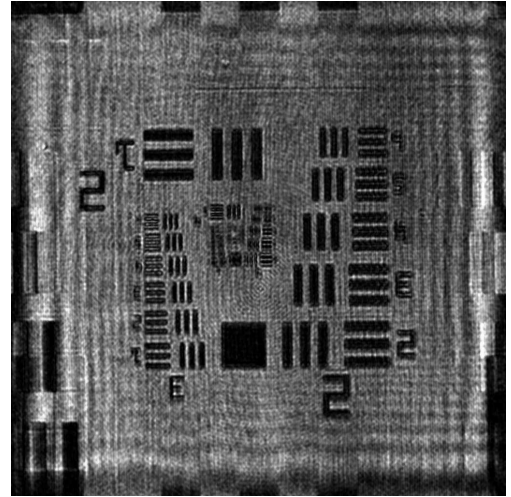
Fig. 2.10 Reconstruction results for the hologram shown in Fig. 2.6(b) by use of the convolution approach: the reconstructed (a) intensity image and (b) phase image with a magnification factor of 6.5; the reconstructed intensity images with magnification factors of (c) 5 and (d) 12, respectively.

To demonstrate the advantages of the convolution approach in the application of digital holography to smaller objects, the reconstruction results for the hologram of the resolution target, originally shown in Fig. 2.8(b), are displayed in Fig. 2.11. However, this time the reconstruction is performed via the convolution approach. As the pixel size in the convolution approach is the same as the pixel size of the sensor chip, rather than depending

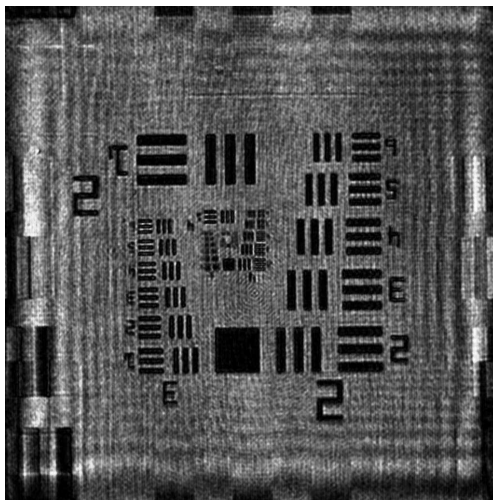
on the recording distance and wavelength, the reconstruction results in Fig. 2.11 can be thought of as an enlargement of the target area. The availability of numerical focusing is also demonstrated here. By changing the distance in the numerical reconstruction, the bars and numbers in the resolution target are either in or out of focus. When the reconstruction distance is equal to the recording distance, i.e. 266mm , the image of the target appears to be sharpest, i.e. fully in focus.



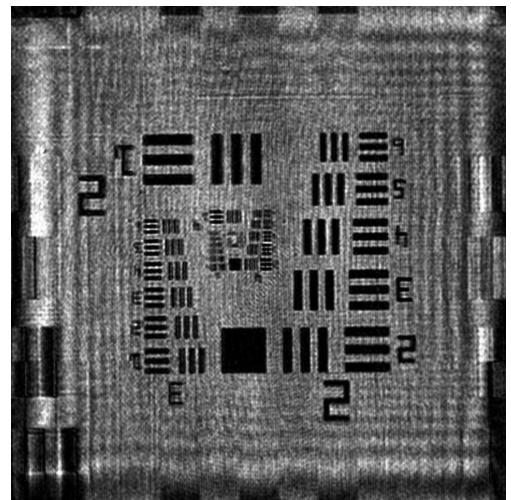
(a)



(b)

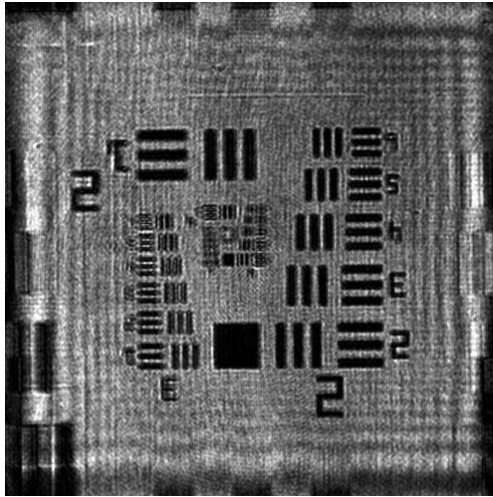


(c)

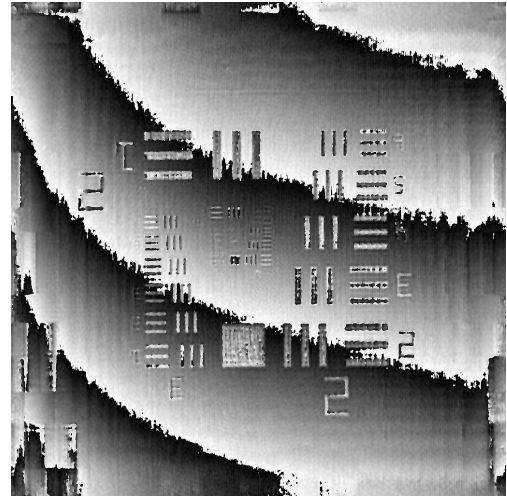


(d)

Fig. 2.11 Reconstruction results for the hologram shown in Fig. 2.8(b) by the convolution approach: the reconstructed intensity-contrast image with a reconstruction distance of (a) 260mm , (b) 264mm , (c) 266mm , (d) 268mm .



(e)



(f)

Fig. 2.11 (Continued) (e) the reconstructed intensity-contrast image with a reconstruction distance of 272mm ; (f) the reconstructed phase-contrast image.

Fig. 2.11(f) is the reconstructed phase-contrast image. The phase of the aluminium bars is quite clear. The large scale fringes in the background area result from the difference between the digital reference wave used in numerical reconstruction and the reference wave in the recording process. Each fringe indicates a jump of 2π of this difference in the phase value. The deviation of these fringes from the straight lines is caused by the phase aberration in the wave to illuminate the resolution target. Both the phase difference and aberration must be removed in quantitative phase-contrast measurement.

2.2.3 Comparison of Fresnel approximation and convolution approach

The Fresnel approximation and the convolution approach are the basic methods of numerical reconstruction in digital holography. Almost all other algorithms are derived from these two algorithms albeit with different improvements and enhancements.

Both methods are practically implemented via the FFT algorithm. In the Fresnel approximation, only a forward FFT is performed. However, two or three FFTs are performed in the convolution approach. This disparity shows the conceptual difference between methods. If the xy plane of the digital hologram is taken as the spatial domain, then the procedure of Eq. (2.24), i.e. Fresnel approximation, gives a result in the frequency domain due to the single Fourier transform. The convolution approach consists of a

forward Fourier transform and an inverse Fourier transform, shown in Eq. (2.37), so the result is still in the spatial domain.

A consequence of this difference is the dissimilarity of the pixel size in the reconstructed images. In the Fresnel approximation, the pixel sizes of the reconstructed image $\Delta\xi$ and $\Delta\eta$ depend on the reconstruction distance and the wavelength used, which is shown in Eq. (2.25). But in the convolution approach, the pixel sizes of the reconstructed image are equal to that of the hologram, shown in Eq. (2.39), and are hence independent of the wavelength and distance.

Fresnel approximation is a “natural scaling” algorithm, as can be seen in Eq. (2.25). The hologram is the aperture of the optical system with a side length of $N\Delta x$. In a diffraction limited system, the half diameter of the Airy disk or speckle diameter restricts the resolution of the system. According to diffraction theory, $\lambda d/N\Delta x$ is equal to the radius of the Airy disk. This sets the image reconstructed by the Fresnel approximation algorithm always to the physical limit of this holographic system. In the convolution approach, the pixel sizes in the reconstructed image are equal to that of the hologram. It would seem that a higher resolution could be achieved if a CCD/CMOS device with a smaller pixel size was used in the recording process. But this resolution is just a numerical value, and the actual resolution is determined by the physical limitations of the optical system as shown in Eq. (2.25).

The inclination factor Q in Eq. (2.1) is set to 1 in digital holography, as illustrated at the beginning of this section. However with the convolution approach, it is possible to vary this parameter and Eq. (2.40) shows that the convolution approach offers the facility to shift a reconstructed area when it is not symmetrical to, or aligned with, the optical axis.

Which reconstruction method to use depends upon the requirements of the specific application that is being considered. Usually, the Fresnel approximation is used more often than the convolution approach when the detected object or surface is bigger than the CCD/CMOS sensor. This is because it runs more quickly, as only a single FFT needs to be carried out. But in particle measurement, microscopy and other applications that detect very small objects, the convolution approach has more advantages and is more accurate than the Fresnel approximation algorithm.

2.2.4 Lensless digital holography (digital Fourier holography)

In the former examples, only plane waves were used as the reference wave. If a spherical wave is used as the reference wave in the holographic system, and the point source of the spherical wave is located in the plane of the object (see Fig. 2.12(a)), it is called lensless holography or Fourier holography. Since the spherical wave can be described as given in Eq. (2.6), the reconstruction formula for the virtual image shown in Eq. (2.26) leads to following equation:

$$\Gamma(\xi, \eta) = C \exp \left[i \frac{\pi}{\lambda d} (\xi^2 + \eta^2) \right] \mathfrak{F}^{-1} [U_h(x, y)] \quad (2.44)$$

where C is a complex constant. Therefore, the reconstruction of a lensless hologram simplifies to the calculation of the Fourier transform of the digital hologram. However, it is not possible to focus on different areas within the object volume in lensless Fourier holography, because the reconstruction distance d does not appear in Eq. (2.44). This means that this technique lacks the ability to perform numerical focusing.

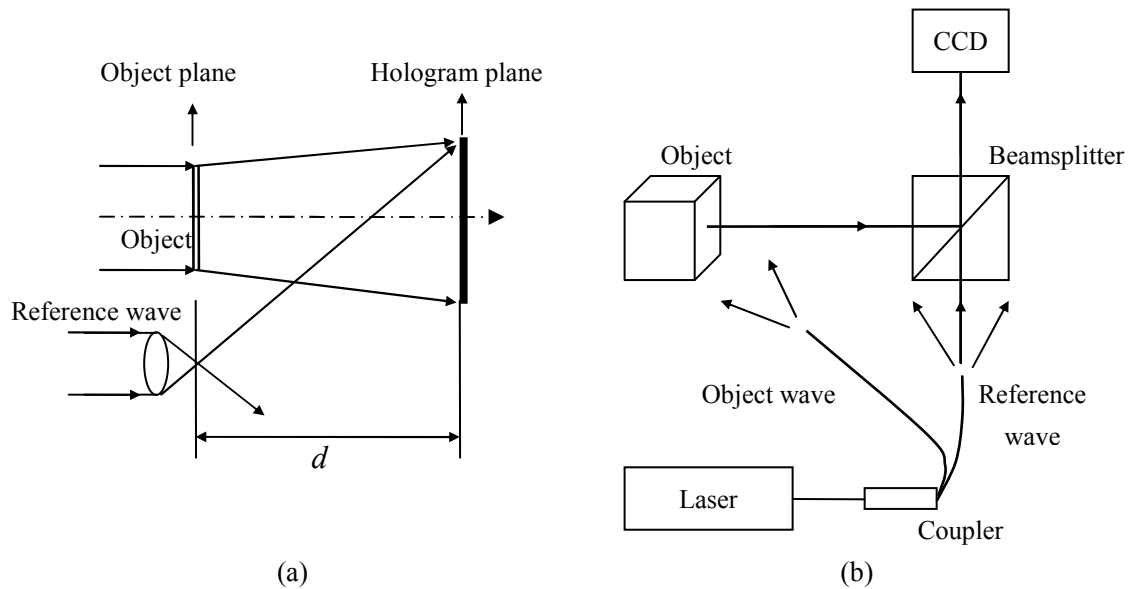
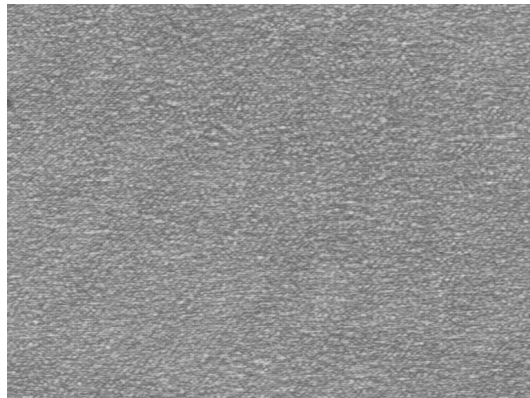


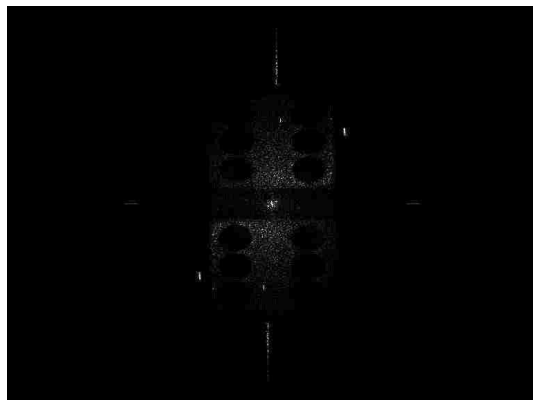
Fig. 2.12 Lensless holography: (a) the principle; (b) the schematic diagram of our experimental system.

Fig. 2.12(b) shows the optical geometry of the experimental system as used here. The use of optical fibres makes the system more flexible, more compact, and more robust. It is also easier to adapt to different types of holography. The same die shown in Fig. 2.6(a) is chosen as the object, because of its small size and the high contrast of the black dots on white background. The light source is a 6mw HeNe laser with a wavelength of 632.8nm. A

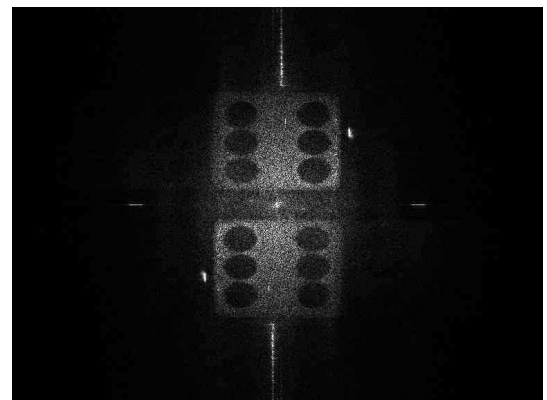
1×2 fibre coupler is used to act as a beamsplitter in order to split the laser beam into two beams with 10% of the output intensity in the reference wave and 90% of the output intensity used to illuminate the object. Usually, the coupling efficiency from laser to single-mode fibre is less than 40%. This means that less than 40% of the output intensity of the laser is transmitted through the fibre. If the object is poorly reflective such as the die, the illumination intensity is not sufficient. Also, the fibre changes the polarization status of the laser beam. So when using low-power lasers, or in the case of applications sensitive to polarization status, optical fibre should be avoided in spite of its advantages of flexibility and miniaturization. The distance between the CCD camera and the object is nominally 720mm. According to the requirement of lensless Fourier holography, the distance between the end of the fibre, where the reference wave exits, and the CCD should also be 720mm – i.e. the same as the distance between the object and detector. A PULNiX TM-6CN CCD camera was used with a Matrox Pulsar frame grabber to capture the digital hologram. The image obtained from this system is of 768×572 pixels and the size of a single pixel is $8.4\mu\text{m}\times 8.4\mu\text{m}$.



(a)



(b)



(c)

Fig. 2.13 Example of lensless digital holography: (a) the lensless Fourier hologram of the die; (b) the reconstructed intensity-contrast image from (a); (c) the reconstructed amplitude-contrast image from (a).

The lensless hologram captured by the CCD camera is shown in Fig. 2.13(a). Applying Eq. (2.44) to this hologram results in the reconstructed intensity- and amplitude-contrast images displayed in Fig. 2.13(b) and (c), respectively. The amplitude-contrast image provides better contrast and more detail on the object is visible. That is because the amplitude of the reconstructed image of the object is much smaller than that of the zero-order term. When it is squared to generate the intensity-contrast image, the intensity of the image of the object is attenuated while the intensity of the zero-order term becomes stronger, which makes the contrast of the image worse than the amplitude-contrast image. There are two dice in the reconstruction results. Both of them are real images which is in accordance with optical lensless holography. The lower one is the primary image and the upper one is the conjugate image, because it is the inverse of the object. The vertical line below the die is the image of the post on which the object is mounted. The light dot in the centre of the image is the zero-order image and the other light dots are caused by optical reflections from the optical components. It is apparent that less than half of the full resolution of the CCD is available in this manner when performing digital holography.

2.2.5 Angular spectrum method

Both the Fresnel approximation and the convolution approach, suffer the same limitation, i.e. that the object under observation must be placed farther away than some minimum distance. If it is placed inside this distance, the spatial frequency of the detector is too low and aliasing occurs. This minimum distance is given by (Hecht, 1998):

$$d_{\min} = \frac{(N\Delta x)^2}{N\lambda} \quad (2.45)$$

where N and Δx are the number and the size of the pixels. However, the angular spectrum method (Kim et al., 2006) is able to overcome this disadvantage. It is comparable with the other methods in terms of computational efficiency but has the potential of higher accuracy.

If the wavefield at the plane $d = 0$ is $U_0(x, y; 0)$, the angular spectrum $A(k_x, k_y; 0)$ at this plane is obtained by taking the Fourier transform:

$$A(k_x, k_y; 0) = \iint U_0(x, y; 0) \exp[-i(k_x x + k_y y)] dx dy \quad (2.46)$$

where k_x and k_y are the corresponding spatial frequencies of x and y . The angular spectrum at the distance d , i.e. $A(k_x, k_y; d)$ is calculated from $A(k_x, k_y; 0)$ as given by:

$$A(k_x, k_y; d) = A(k_x, k_y; 0) \exp(ik_z d) \quad (2.47)$$

where $k_z = \sqrt{k^2 - k_x^2 - k_y^2}$. The reconstructed complex wavefield at any plane perpendicular to the propagating z axis is found by

$$U(\xi, \eta; d) = \iint A(k_x, k_y; d) \exp[i(k_x \xi + k_y \eta)] dk_x dk_y = \mathfrak{T}^{-1}[\mathfrak{T}(U_0) \exp(ik_z d)] \quad (2.48)$$

The resolution of the reconstructed images from the angular spectrum method is the same as that in the hologram plane, which means that the pixel size does not vary with changes of wavelength or reconstruction distance. In Fig. 2.14, the results obtained from using this method clearly show the significant advantage of the angular spectrum method in calculating wave fields near the hologram plane. The object in Fig. 2.14 was part of the surface of a coin with an area of $2.62\text{mm} \times 2.62\text{mm}$. The reconstructed distance is 3.9mm . The wavelength used to record Fig. 2.14(a) was 575nm .

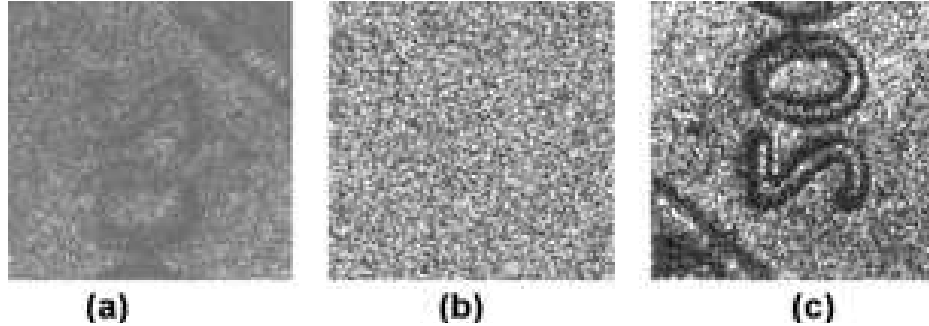


Fig. 2.14 (a) Hologram of a penny; reconstruction from (b) convolution approach and (c) the angular spectrum method. (Reprinted with permission from Yu and Kim (2005). Copyright (2005), Optical Society of America.)

2.2.6 Phase retrieval methods

The methods mentioned above quite literally translate the physical process of the optical reconstruction of conventional holograms into a numerical analogue. However, some researchers have gone beyond simple mimicking of conventional holographic practice and have attempted to gain access to the additional information numerical reconstruction can offer. In particular they have applied the concept of phase retrieval to the numerical reconstruction in digital holography.

Liebling et al (2004) perform a nonlinear change of variables so that the reconstruction may be performed by use of a method that is reminiscent of phase-shifting techniques. The algorithm is based on a local least-squares estimation of the amplitude and phase by assuming an *a priori* model of the phase of the reference wave. Once the complex object wave is recovered in the acquisition plane, the wave is back-propagated to restore a focused image using a digital implementation of the Fresnel transform.

Liu and Scott (1987) proposed an approach to implement numerical reconstruction that uses the Gerchberg-Saxton phase-retrieval algorithm. Zhang et al (2003) introduced the Yang-Gu phase-retrieval algorithm to the numerical reconstruction of in-line digital holograms. Iterations are implemented in both methods until convergence is achieved when the difference between the hologram obtained by the iteration and the real one captured by camera reaches a minimum value. The Yang-Gu algorithm takes more time to process than the Gerchberg-Saxton algorithm because it has two iteration loops. But it can be easily expanded to multiwavelength and multiplane systems and is not sensitive to correct choice of initial value.

The method proposed by Zheng et al (2005) is based on the processing of a series of holograms of the same object recorded with different distances. What can be recorded on the CCD/CMOS device is the intensity distribution, and the phase information is lost completely. So the direct application of the Fresnel approximation to the hologram will produce both the original image and its twin image. In order to solve this problem, the phase of the wave field must be retrieved. For the hologram I_1 recorded at a distance d_0 from the object, a propagation of this pattern I_1 from d_0 to $d_0 + \Delta d$ using the Fresnel approximation is calculated obtaining a complex wave field distribution A'_2 . Then keeping the phase distribution of the field A'_2 unchanged and using the square root of the hologram I_2 as an amplitude distribution, thus a new wave field A_2 is composed. Then a propagation of this new wave field A_2 is performed from $d_0 + \Delta d$ to $d_0 + 2\Delta d$. After determination of the phase and amplitude of the wavefront at a given plane, it is possible to recover the image of the object by using the inverse Fresnel transform. The phase distribution was revised again and again by using the correct intensity information provided by multiple holograms, and the real phase distribution can be obtained in this manner. It is very similar to the algorithm for retrieving phase information from two intensity distributions, such as the Gerchberg-Saxton algorithm and the Yang-Gu algorithm,

but the use of multiple holograms rather than two intensity distributions means that more information is available to recover the phase.

2.2.7 Other algorithms for numerical reconstruction

Zhang et al (2005) proposed a systematic approach from the perspective of “generalized sampling theory”. An approximation of the continuous complex amplitude at the CCD/CMOS sensor can be synthesized from a set of basis functions, with the recorded samples as weights. By adopting different basis functions and different formulas for describing the diffraction process, an optimal reconstruction algorithm can be developed for various recording conditions and different diffraction characteristics of the object.

Yu and Cai (2001) proposed a method to numerically extract the object information from the fringe pattern of the hologram. Then an iterative algorithm is used to imitate an imaging system by focusing on different layers of the object; and by operation in both the spatial domain and the frequency domain, the algorithm produces a series of two-dimensional layer images. The object is finally reconstructed layer by layer subject to a constraint condition which must be satisfied.

Liebling et al (2003) constructed a new wavelet basis, Fresnelets, for the processing and reconstruction of digital holograms by taking advantage of the mathematical properties of the Fresnel transform. It allows reconstruction at different user-specified and wavelength independent scales. The reconstructions at a coarse scale allow for optimal filtering of the zero-order and the twin image and also result in less noisy images.

2.3 Conclusions

In this chapter, various algorithms for the numerical reconstruction of digital holograms have been presented and discussed. The main focus has been on the Fresnel approximation and the convolution approach as they are the direct derivatives from scalar diffraction theory and most other algorithms are connected with them at least to some degree. The Fresnel approximation is the most cost-effective algorithm due to its use of only one FFT. For the measurement of macroscopic objects at longer distances, say some tens of millimetres, the Fresnel approximation is the best option. But for microscopic objects, with their rather small size, the convolution approach shows its advantages as demonstrated in

Fig. 2.11. The angular spectrum algorithm provides more accurate reconstruction results than both the Fresnel approximation and the convolution approach, which can be seen clearly in Fig. 2.14. But for the purpose of this thesis, i.e. measuring macroscopic objects, the recording distance between the object plane and the CCD/CMOS plane is sufficient to ensure that the Fresnel approximation is equivalent to the angular spectrum algorithm, but with significantly greater simplicity. Phase retrieval methods can avoid the reconstructed image being severely corrupted by the zero-order term and the twin-image. They intrinsically remove these terms without the need for any pre- or post-processing. However they are complex procedures and the implementations of phase retrieval methods are based on iteration loops, which means that they are more time-consuming than other algorithms. Other algorithms have been developed for some specific applications, or were used only in initial development stages. They have not been used as widely as the Fresnel approximation, the convolution approach and the angular spectrum algorithm.

In this research both computer simulations and experiments were performed in order to compare the Fresnel approximation algorithm and the convolution approach. Through the results shown in this chapter, the Fresnel approximation algorithm can be seen to provide a better solution to carry out numerical reconstruction for macroscopic objects. The IDL codes for performing these two algorithms and the computer simulations for the formation of the digital hologram are enclosed in the Appendix.

In the latter part of this thesis, it may be assumed that, unless otherwise indicated, the Fresnel approximation algorithm is always used to conduct numerical reconstruction of the digital holograms.

References:

- CUCHE, E., BOVILACQUA, F. & DEPEURSIAGE, C. (1999) Digital holography for quantitative phase-contrast imaging. *Optics Letters*, 24, 291-3.
- DEMETRAKOPOULOS, T. H. & MITTRA, R. (1974) Digital and optical reconstruction of images from suboptical diffraction patterns. *Applied Optics*, 13, 665-670.
- GOODMAN, J. W. (2005) *Introduction to Fourier Optics (Third Edition)*, Englewood, Colorado, Roberts & Company.
- HECHT, E. (1998) *Optics*, Addison Wesley Longman, Inc., p 485.
- KIM, M. K., YU, L. & MANN, C. J. (2006) Interference techniques in digital holography. *Journal of Optics A: Pure and Applied Optics*, 8, 518-523.
- KREIS, T. M. & JUPTNER, W. (1997) Principles of digital holography. IN JUPTNER, W. & OSTEN, W. (Eds.) *Fringe '97, Proceeding of 3rd International Workshop on Automatic Processing of Fringe Patterns*. Akademie, Berlin.
- LIEBLING, M., BLU, T. & UNSER, M. (2003) Fresnelets: New multiresolution wavelet bases for digital holography. *IEEE Transactions on Image Processing*, 12, 29-43.
- LIEBLING, M., BLU, T. & UNSER, M. (2004) Complex-wave retrieval from a single off-axis hologram. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 21, 367-377.
- LIU, G. & SCOTT, P. D. (1987) Phase retrieval and twin-image elimination for in-line Fresnel holograms. *Journal Of The Optical Society Of America A*, 4, 159-165.
- SCHNARS, U. & JUPTNER, W. (2005) *Digital holography - digital hologram recording, numerical reconstruction, and related techniques*, Springer.
- SCHNARS, U. & JUPTNER, W. P. O. (1994) Digital recording and reconstruction of holograms in hologram interferometry and shearography. *Applied Optics*, 33, 4373-7.
- YAROSLAVSKY, L. P. & MERZLYAKOV, N. S. (1980) *Methods of digital holography*, New York, Consultants Bureau.
- YU, L. & CAI, L. (2001) Iterative algorithm with a constraint condition for numerical reconstruction of a three-dimensional object from its hologram. *Journal of the Optical Society of America A (Optics, Image Science and Vision)*, 18, 1033-45.

- YU, L. & KIM, M. K. (2005) Wavelength-scanning digital interference holography for tomographic three-dimensional imaging by use of the angular spectrum method. *Optics Letters*, 30, 2092-2094.
- ZHANG, F., YAMAGUCHI, I. & YAROSLAVSKY, L. (2005) Systematic approach for the design of reconstruction algorithm in digital holography. Beijing, China, International Society for Optical Engineering, Bellingham, WA 98227-0010, United States.
- ZHANG, Y., PEDRINI, G., OSTEN, W. & TIZIANI, H. J. (2003) Image reconstruction for in-line holography with the Yang-Gu algorithm. *Applied Optics*, 42, 6452-6457.
- ZHENG, D.-X., ZHANG, Y., SHEN, J.-L., ZHANG, C.-L. & PEDRINI, G. (2005) Wave field reconstruction from a hologram sequence. *Optics Communications*, 249, 73-77.

3 Suppression of the Zero-Order and Twin-Image Terms

3.1 Introduction

Suppose a hologram $U_h(x, y)$ is generated by the interference of an object wave $U_o(x, y)$ and a reference wave $U_r(x, y)$. The hologram reconstruction can be expressed in Eq. (3.1).

$$U_r(x, y)U_h(x, y) = U_r(x, y)\left(|U_r(x, y)|^2 + |U_o(x, y)|^2\right) + U_o(x, y)|U_r(x, y)|^2 + U_o^*(x, y)U_r^2(x, y) \quad (3.1)$$

where $*$ stands for the complex conjugate.

The coordinate system for digital holography is shown in Fig. 3.1.

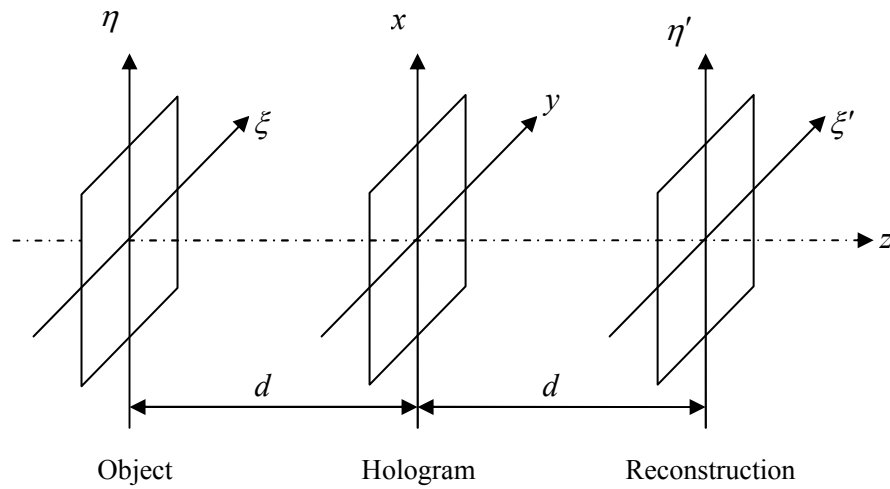


Fig. 3.1 The coordinate system of digital holography

As discussed in the previous chapter, there are three terms in the reconstructed field. The first term is known as zero-order term (or dc term) which consists of the intensity of the reference wave superposed with that of the object wave. The second and the third terms are the reconstructed object wave and conjugate object wave, respectively. The conjugate object wave is the twin-image term. In Gabor's in-line holography, these three terms superpose with each other and cannot be observed separately. Off-axis holography invented by Leith and Upatneiks (1962) introduced a spatial carrier so that the zero-order

term, the image of the object and the twin image can be physically separated from each other. This is also widely used in digital holography. Fig. 3.2 displays a typical off-axis hologram and its reconstructed amplitude image.

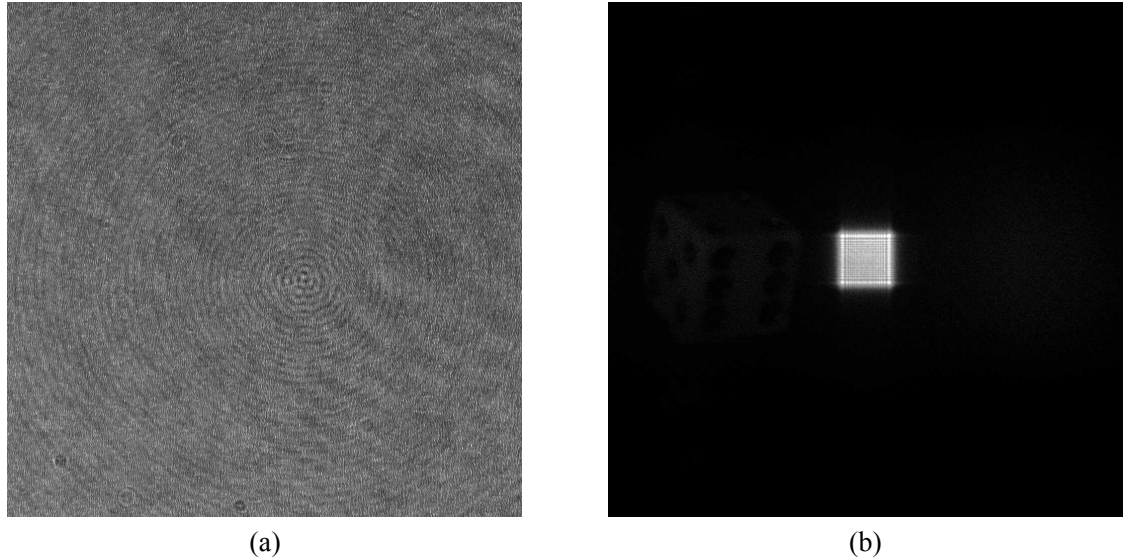


Fig. 3.2 (a) A typical digital hologram of a die; (b) the reconstructed amplitude of hologram in (a).

The very bright square in the centre of Fig. 3.2(b) is the zero-order term. The image of the die is quite faint and can be seen on the left and the twin-image is on the right of the whole picture. It is because the zero-order term is so intense that the image of the die displays very faintly and the twin image is totally submerged in the background. Due to the limited resolution and dynamic range of CCD/CMOS sensors, the zero-order term and twin image must be suppressed in order to display the desired image normally so that it has adequate brightness and contrast levels. This is the usual case when displaying the reconstruction results, rather than the raw unprocessed image as shown in Fig. 3.2(b). Many approaches have been proposed to suppress the influence of the zero-order term and twin image. They can be classified into two categories according to their foundation: digital image processing methods and experimental methods. In this chapter, we will discuss different methods to suppress both the zero-order term and the twin image, for in-line digital holography and off-axis digital holography.

3.2 Suppression of zero-order term and twin image in off-axis holograms

3.2.1 Methods based on digital image processing

The physical meaning of zero-order term is that it is the zero-order diffraction of the reference wave, or equivalently the projection of the illuminated CCD/CMOS array. As illustrated in the previous chapter on numerical reconstruction of digital holograms, the Fresnel transform is the Fourier transform of a product of the hologram and the chirp function, if the plane reference wave is applied according to Eq. (2.20). A chirp function indicates a signal in which the frequency changes with time enabling the characterisation of frequency response by the application of a single signal. The chirp function as used in digital holography will be discussed later in this section. It can also be treated as the convolution of the hologram and the chirp function in Eq. (2.34). The Fourier transform of the hologram is trimodal with a peak at the spatial frequency (0,0). A schematic of the spatial spectrum of an off-axis digital hologram is shown in Fig. 3.3. The zero-order term is in the centre and the two sidelobes indicate the image of the object and its twin image. Fig. 3.4 is the spatial spectrum of the hologram that was shown in Fig. 3.2(a).

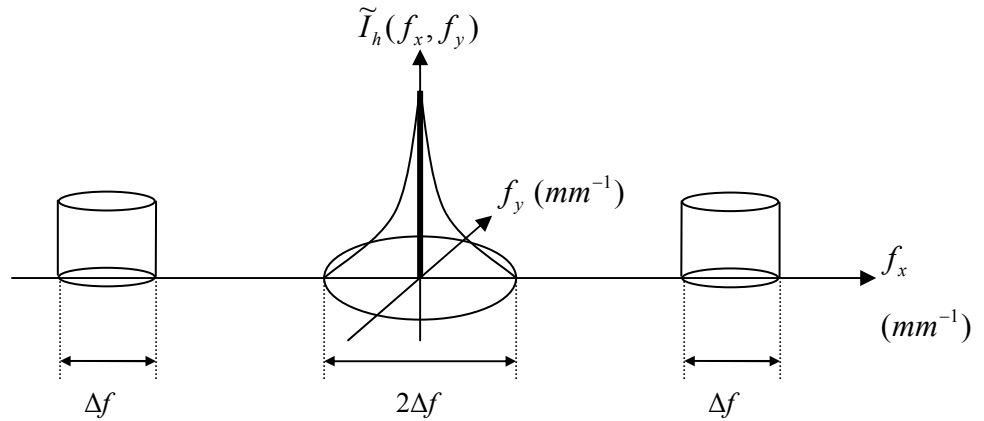


Fig. 3.3 A schematic of the spatial spectrum of a digital hologram (redrawn from (Liu et al., 2002))

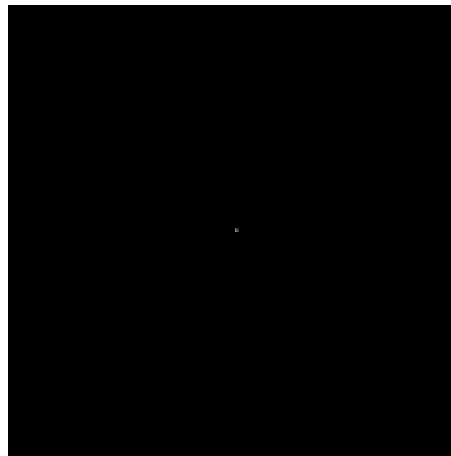


Fig. 3.4 The spatial spectrum of the digital hologram shown in Fig. 3.2(a)

In Fig. 3.4, only the zero-order term in the centre of the spectrum is displayed. It shows that the majority of the intensity is constrained in the zero-order term so that the information in other frequency is too weak to be seen in the spatial spectrum of this off-axis hologram. This explains what is shown in Fig. 3.2(b).

The zero-order term of the spatial spectrum of the hologram is calculated by the following relation:

$$\tilde{U}_h(0,0) = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} U_h(k\Delta x, l\Delta y) \quad (3.2)$$

This can be modelled by a Dirac delta function in the frequency domain. So the zero-order term of the Fresnel transform of the digital hologram can be considered to be the zero-order term of the Fourier transform of the digital hologram convolved with the Fourier transform of the two-dimensional chirp function. Since a Dirac delta function is used to represent the former, the zero-order term of the whole Fresnel transform is the Fourier transform of the chirp function:

$$\exp\left(-i\pi \frac{x^2 + y^2}{\lambda d}\right) = \exp\left(-\frac{i\pi}{\lambda d} x^2\right) \exp\left(-\frac{i\pi}{\lambda d} y^2\right) \quad (3.3)$$

λ is the wavelength used in the reconstruction, and d is the distance between the hologram plane and the reconstructed image plane. Considering the chirp function is separable in Eq. (3.3), it is sufficient to investigate the one-dimensional chirp function. Fig. 3.5(a) shows the real part of a one-dimensional finite chirp function. The one-dimensional spectrum of this finite chirp function $\exp(-i\pi x^2/(\lambda d))$ is shown in Fig. 3.5(b). The spectrum of the chirp function is almost flat over a certain region and almost zero outside that region.

According to the analysis by Kreis and Juptner (1997), the width of the zero-order term is of $N^2\Delta x^2/(\lambda d)$ pixels in the x axis and $N^2\Delta y^2/(\lambda d)$ pixels in the y axis, where Δx and Δy are the pixel spacings in x and y axes, respectively. The width of the zero-order term increases with an increase of the pixel size and also with the number of pixels in the CCD sensor, while it decreases when the distance and wavelength increase.

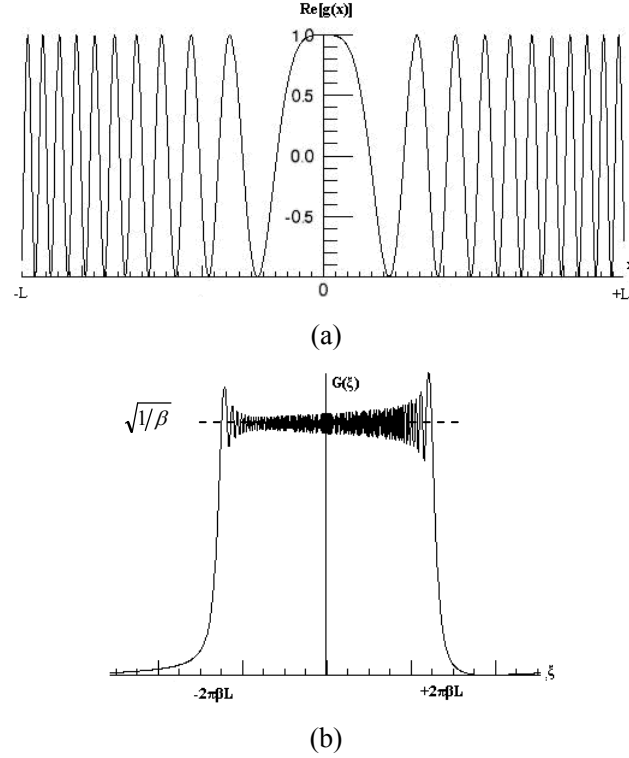


Fig. 3.5 (a) The finite chirp function; (b) its spectrum (L is the half of the side length of the hologram, d and λ are the distance and the wavelength used in the reconstruction, respectively, and $\beta = 1/(\lambda d)$.)

Since the zero-order term of the hologram spectrum is the sum of the intensities of all pixels according to Eq. (3.2), Kreis and Juptner (1997) proposed subtracting the average intensity of the hologram from the intensity value of every pixel. In this manner the zero-order term in the spectrum of the hologram is zero. The average intensity value is calculated by

$$U_{hm} = \frac{1}{NM} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} U_h(k\Delta x, l\Delta y) \quad (3.4)$$

The modified hologram

$$U'_h(k\Delta x, l\Delta y) = U_h(k\Delta x, l\Delta y) - U_{hm}, \quad k = 0, \dots, N-1, \quad l = 0, \dots, M-1 \quad (3.5)$$

is processed by a normal numerical reconstruction algorithm in order to obtain the object image and twin image, but with the zero-order term effectively suppressed. There must be some negative values in the modified hologram. This cannot be achieved by optical methods, but for digital image processing it is feasible. The whole hologram is downshifted in its intensity with an unchanged relationship between the intensities of individual pixels. The reconstruction results for the hologram shown in Fig. 3.2(a), when

processed by this method, are presented in Fig. 3.6(b). Although the zero-order term in the spatial spectrum is quite strong, as shown in Fig. 3.6(a), it is suppressed and the two sidelobes can now be seen that were not visible in Fig. 3.4. The effect of the suppression of the zero-order term is verified in Fig. 3.6(b).

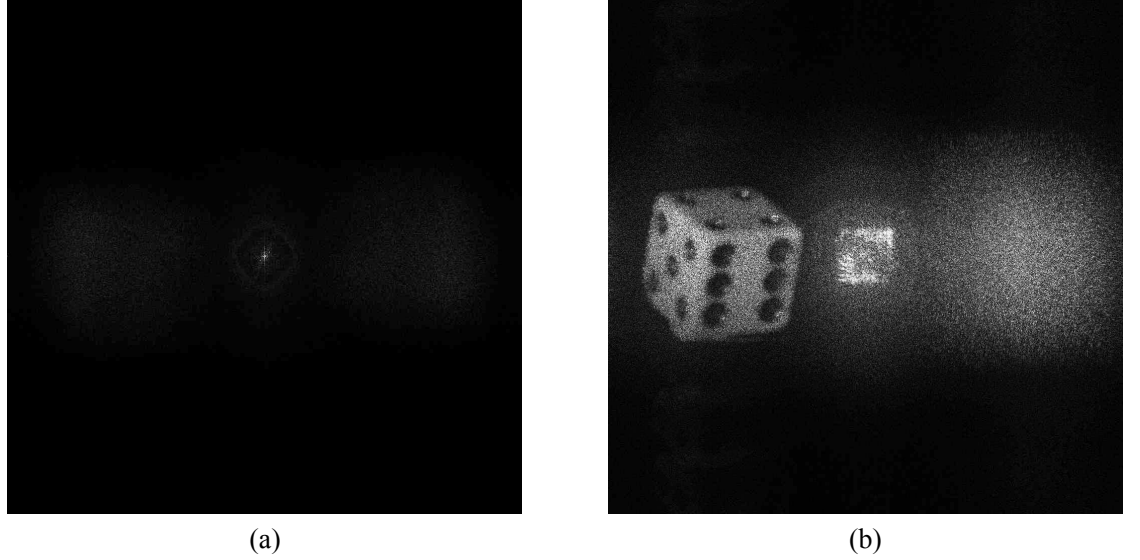


Fig. 3.6 The zero-order term suppression by subtracting the mean pixel value from the hologram before numerical reconstruction: (a) the spatial spectrum and (b) the reconstructed amplitude-contrast image of the processed hologram.

This method can also be interpreted as a high-pass filtering operation, with a low cut-off frequency equal to the smallest nonzero frequency (Kreis and Juptner, 1997). Various high-pass filters are used and good results have been realized by subtracting the low-pass filter result of the average over each 3×3 pixel neighbourhood from the original digital hologram.

$$\begin{aligned}
 U'_h(k, l) = & U_h(k, l) - \frac{1}{9} [U_h(k-1, l-1) + U_h(k-1, l) + U_h(k-1, l+1) \\
 & + U_h(k, l-1) + U_h(k, l) + U_h(k, l+1) + U_h(k+1, l-1) + U_h(k+1, l) + U_h(k+1, l+1)] \\
 & k = 1, \dots, N-2, \quad l = 1, \dots, M-2
 \end{aligned} \tag{3.6}$$

The factors Δx and Δy in the pixel arguments have been omitted for convenience. The results provided by implementing this method can be found in Fig. 3.7. The hologram of the die shown in Fig. 3.2(a) is used for this example. In the spatial spectrum of the high-pass filtered hologram, the zero-order term has disappeared completely. Therefore, in the reconstructed amplitude-contrast image, the central bright square is not present. However, the filtering process caused the amplitude of the object at locations near to the position of the zero-order term to decrease.

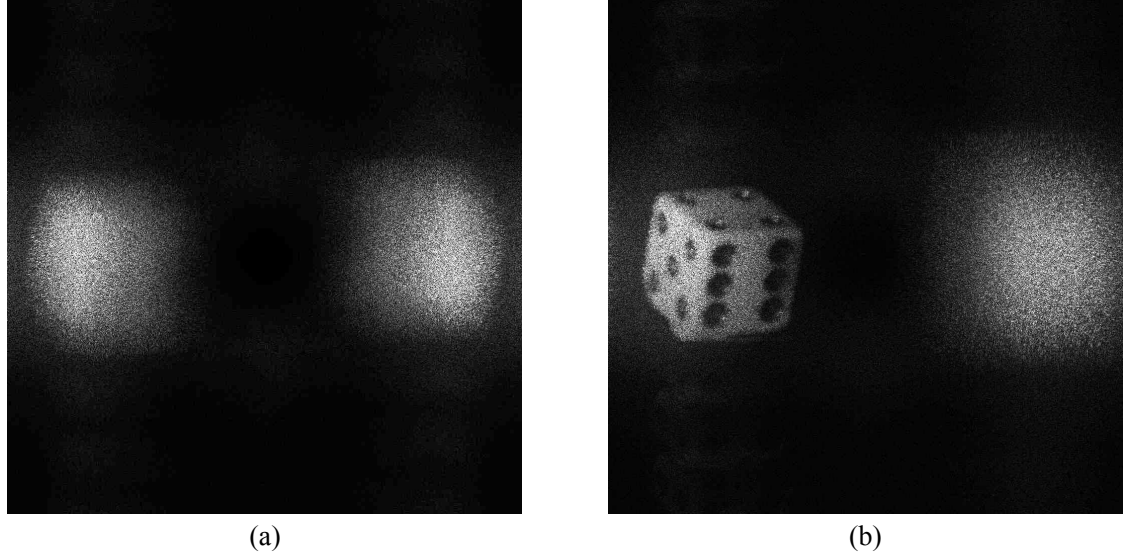


Fig. 3.7 The zero-order term suppression by high-pass filtering the hologram before numerical reconstruction: (a) the spatial spectrum and (b) the reconstructed amplitude-contrast image of the processed hologram.

Liu et al. (2002) used the Laplacian of the off-axis hologram, instead of the hologram itself, for numerical reconstruction by computer. The negative Laplacian of the hologram intensity is defined as

$$L_h(x, y) = -\nabla^2 U_h(x, y) = -\left(\frac{\partial^2 U_h(x, y)}{\partial x^2} + \frac{\partial^2 U_h(x, y)}{\partial y^2} \right) \quad (3.7)$$

The discrete form of Eq. (3.7) can be written as

$$\begin{aligned} L_h(k, l) &= -\nabla^2 U_h(k, l) \\ &\approx -\frac{1}{\Delta x^2} [U_h(k+1, l) + U_h(k-1, l) + U_h(k, l+1) + U_h(k, l-1) - 4U_h(k, l)] \end{aligned} \quad (3.8)$$

Performing a Fourier transformation on both sides of Eq. (3.8) and simplifying it, gives us:

$$\begin{aligned} \tilde{L}_h(f_x, f_y) &= -\frac{2}{\Delta x^2} [\cos(2\pi\Delta f_x) + \cos(2\pi\Delta f_y) - 2] \tilde{U}_h(f_x, f_y) \\ &= 4 \left[\frac{\sin^2(\pi\Delta f_x) + \sin^2(\pi\Delta f_y)}{\Delta x^2} \right] \tilde{U}_h(f_x, f_y) \end{aligned} \quad (3.9)$$

where $\tilde{L}_h(f_x, f_y)$ and $\tilde{U}_h(f_x, f_y)$ are the Fourier transforms of $L_h(x, y)$ and $U_h(x, y)$, respectively, Δf_x and Δf_y are the spacings of the f_x and f_y axes in spatial frequency domain. Also due to the separability of this equation, we need only to investigate the one-dimensional spectrum $\tilde{L}_h(f_x)$, which is shown schematically in Fig. 3.8. The

spectrum of the zero-order term is efficiently suppressed, and the spectra of the images are almost intact except for a small change in intensity. Therefore, it can be seen that the quality of the reconstructed image can be significantly improved when the Laplacian of the detected hologram is used for reconstruction instead of the hologram itself.

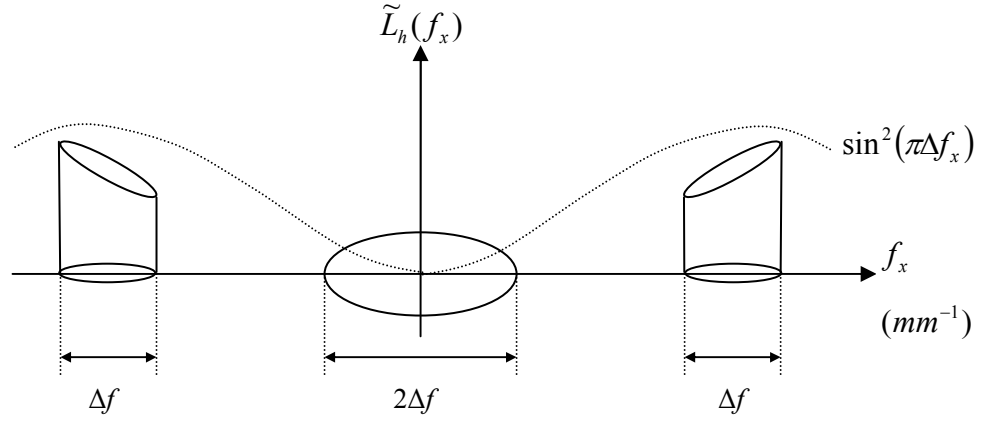


Fig. 3.8 The schematic of the spatial spectrum of the Laplacian of the digital hologram (redrawn from (Liu et al., 2002))

In Fig. 3.8 the modification in the spectrum of the image of the object and the twin image has been introduced by using the Laplacian of the digital hologram for numerical reconstruction. This means that the reconstructed image will be modified to some extent. However, the maximum spectral width of the image of the object is much smaller than the period of the function $\sin^2(\pi\Delta f_x)$. So there is only a small change in the intensity of both of the images. On the other hand, $\sin^2(\pi\Delta f_x)$ is symmetrical and has the same effect on both images: weakening their lower-frequency components and strengthening their higher-frequency components. According to Fourier imaging principles, this kind of modification only results in an alteration of the image contrast. Using this method does affect the image quality of both images, but it is acceptable in comparison with the significant improvement in image quality due to the elimination of the zero-order term. An example of using this method is shown in Fig. 3.9. The hologram of the die originally shown in Fig. 3.2(a) is used to verify this method. The results look quite similar to those obtained from the high-pass filtering method presented previously in Fig. 3.7. The zero-order term is completely removed and the information of the die is preserved and enhanced. Meanwhile, the signal-to-noise ratios of Fig. 3.7(b) and Fig. 3.9(b) are similar, being 1.14 and 1.18, respectively. The only difference between the results from the

Laplacian method and the high-pass filtering method lies in the fact that Fig. 3.9(b) displays more ambient noise around the image of the die and its twin image.

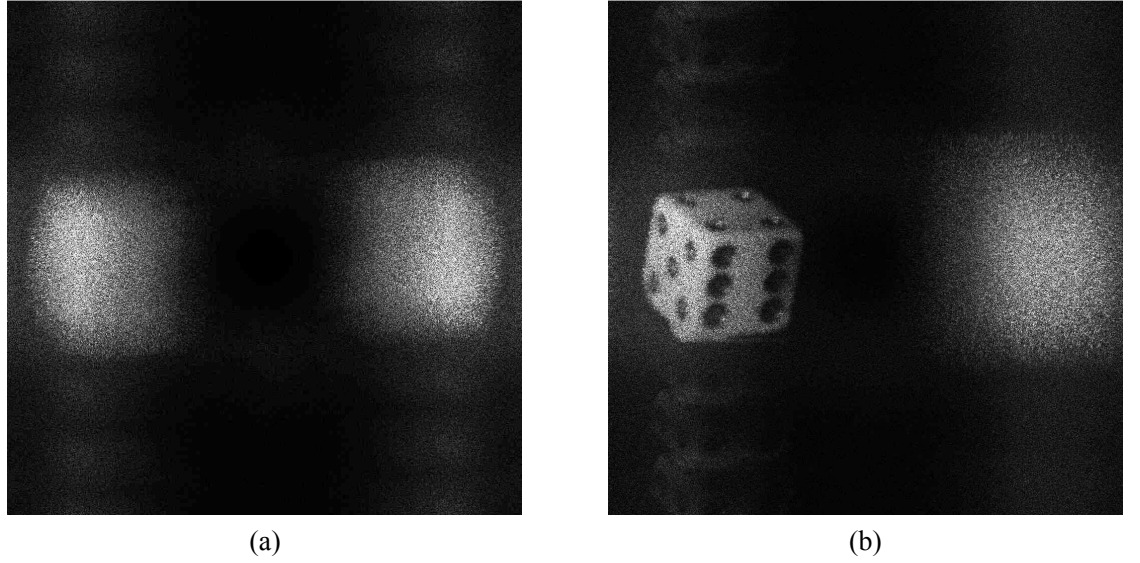


Fig. 3.9 The zero-order term suppression by using the Laplacian of the hologram: (a) the spatial spectrum and (b) the reconstructed amplitude-contrast image of the processed hologram.

Recently, a new numerical method to suppress the zero-order image has been proposed (Chen et al., 2007). Suppose the object wave to be detected and reconstructed is U_o and the reference wave is U_r . The hologram generated by these two waves can be written as

$$U_h = |U_o + U_r|^2 = |U_o|^2 + |U_r|^2 + U_o^* U_r + U_o U_r^* \quad (3.10)$$

The first two terms on the right-hand side of Eq. (3.10) consist of the zero-order image in the reconstructed result, as illustrated in the previous sections, where $|U_o|^2$ is the intensity of the object wave and $|U_r|^2$ is the intensity of the reference wave. These two terms must be eliminated in order to suppress the zero-order image. Usually the amplitude and the phase of the reference wave are known. Therefore subtracting $|U_r|^2$ from the hologram in Eq. (3.10) and squaring the result yields

$$\begin{aligned} (U_h - |U_r|^2)^2 &= (|U_o|^2 + U_o^* U_r + U_o U_r^*)^2 \\ &= |U_o|^2 |U_o|^2 + 2|U_o|^2 (U_o^* U_r + U_o U_r^*) + (U_o^* U_r + U_o U_r^*)^2 \\ &= |U_o|^2 (2U_h - |U_o|^2) + (U_o^* U_r)^2 + (U_o U_r^*)^2 \end{aligned} \quad (3.11)$$

Since $(U_o U_r^*)^2$ and $(U_o^* U_r)^2$ are mutually conjugate, the sum of them is real, because the imaginary part cancels out. So the sum can be substituted by $2|U_o U_r^*|^2$ and Eq. (3.11) can be rewritten as

$$|U_o|^2 = \frac{(U_h - |U_r|^2)^2}{2(U_h - |U_o|^2/2 + |U_r|^2)} \propto \frac{(U_h - |U_r|^2)^2}{U_h + |U_r|^2} \quad (3.12)$$

It is assumed that the influence of $|U_o|^2/2$ can be neglected, because the intensity of the reference wave is larger than that of the object wave. Finally, the first two terms on the right-hand side of Eq. (3.10) are removed as given by:

$$U_o U_r^* + U_o^* U_r = U_h - |U_r|^2 - \frac{(U_h - |U_r|^2)^2}{U_h + |U_r|^2} \quad (3.13)$$

To use Eq. (3.13) to suppress zero-order image, $|U_r|^2$ must be known so that it can be numerically generated by computer. Fig. 3.10 shows the simulation results for this method.

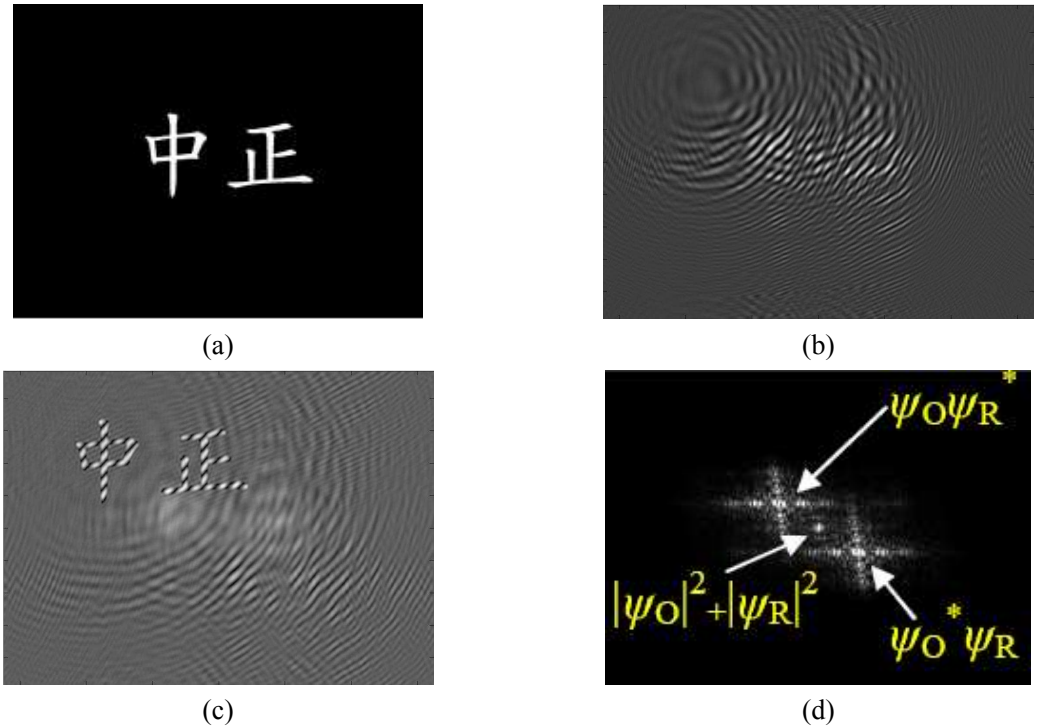


Fig. 3.10 Simulation results: (a) source image; (b) digital hologram; (c) reconstructed image; (d) Fourier power spectrum of (c).



Fig. 3.10 (Continued) (e) reconstructed image obtained by novel suppression approach; (f) Fourier power spectrum of (e). (Reprinted with permission from Chen et al. (2007). Copyright(2007), Optical Society of America)

In this simulation, the recording wavelength is $632.8nm$. The distance between the object plane and hologram plane is $286mm$. Fig. 3.10(a) is the input image to generate the object wave. By the interference with the off-axis plane reference wave of the wave vector $k_x = 3380mm^{-1}$ and $k_y = 3240mm^{-1}$, the calculated hologram is shown in Fig. 3.10(b). Fig. 3.10(c) presents the reconstructed image obtained by performing the numerical propagation of $286mm$ with the digital hologram. The reconstructed image is superposed on the zero-order image so that it cannot be resolved. Fig. 3.10(d) shows the calculated Fourier spectrum of Fig. 3.10(c). Fig. 3.10(e) shows the reconstructed image by performing the method as stated in Eq. (3.13). The object image is resolved effectively because the zero-order term is deeply suppressed, which is clearly shown in Fig. 3.10(f).

In practical terms, some questions must be raised about performing this method. The intensity distribution of the reference wave has to be generated numerically in order to help suppress the zero-order term. In the experimental work published in (Chen et al., 2007), it stated that they achieved this by “multiplying the proportion of the intensity ratio (between the object wave and the reference wave) with an operation array”. The meaning of this sentence is so ambiguous that it is not clear exactly how to carry out this process. Meanwhile, they only applied this new approach to transmission digital holography, in which it is easy to measure the intensity ratio of the object wave and the reference wave. But for applications that deal with diffusely reflective objects, it is quite difficult to measure this intensity ratio accurately.

For the methods listed above, the zero-order term is effectively suppressed, while the twin image is either intact or has only a small or limited change. However, the existence of the twin image adversely affects the quality of the desired image. A method based on spatial

filtering proposed by Cuche et al. (2000) can deal with this problem. In off-axis holography, the object beam $U_o(x, y)$ and the reference beam $U_r(x, y)$ arrive at the hologram plane from two different directions. Therefore the different terms will vary at different spatial frequencies and, as a consequence, will propagate along the two separate directions during the reconstruction. For example, if we assume a reference wave of the form $U_r(x, y) = E_r \exp(ikx \sin \theta)$, where θ is the angle between the object wave and the reference wave, the hologram intensity becomes

$$U_h(x, y) = E_r^2 + I_o(x, y) + \exp(-ikx \sin \theta) E_r U_o(x, y) + \exp(ikx \sin \theta) E_r U_o^*(x, y) \quad (3.14)$$

The phase factor $\exp(-ikx \sin \theta)$ in the third term and the phase factor $\exp(ikx \sin \theta)$ in the fourth term indicate the deflection of the reconstructed images from $U_o(x, y)$, the wave used to illuminate the hologram.

Considering the Fourier transform of the hologram, the influence of the two phase factors $\exp(\pm ikx \sin \theta)$ can be interpreted as a translation of the spatial frequencies associated with the real and the virtual images. Assume that $U_o(x, y)$ is in normal incidence, the spatial frequencies of the zero-order diffraction are located in the centre of the Fourier plane, and the spatial frequencies of the interference terms vary at different carrier frequencies, which are located symmetrically with respect to the centre of the Fourier plane: $-k \sin \theta / 2\pi$ for the virtual image and $k \sin \theta / 2\pi$ for the real image. This suggests that the different terms of the reconstructed wave front can be spatially filtered. From the point of view of signal processing it is interesting to realize that the off-axis geometry introduces a spatial modulation of the interference terms. This task could be achieved optically during a conventional optical reconstruction with two lenses in a $4f$ configuration with an appropriate mask inserted in the Fourier plane of the first lens. However in digital holography this procedure can be performed digitally by multiplication of the computed Fourier transform of the hologram with a numerically defined mask.

After performing the forward Fourier transform with an off-axis hologram, the different interference terms produce well-separated contributions. Spatial frequencies corresponding to the zero-order diffraction are located in the centre of the image, which includes the contribution of the reference intensity and the object intensity. The spatial frequencies of the third and fourth terms in Eq. (3.10) are located symmetrically with respect to the centre

of the image. This Fourier-transformed hologram is then multiplied with a mask that eliminates all spatial frequencies except those of the interference terms. Then the inverse Fourier transform is applied in order to generate a filtered hologram. It is this filtered hologram, instead of the original hologram, that is used in the numerical reconstruction. The mask used for spatial filtering consists of two transparent windows centered around the carrier frequencies of the interference terms. Windows with smoothed transmission profiles such as Gaussian, Hamming, Butterworth, or Tukey windows can be used to avoid the occurrence of high-frequency fluctuations in the reconstructed images.

To eliminate the twin image, an asymmetrical mask may be used composed of only one transparent window centred on the carrier frequency of one of the interference terms. Since the hologram is a real function, its Fourier transform is symmetric with its complex conjugate. This symmetry is preserved if the mask used for spatial filtering is also symmetric. Thus the inverse Fourier transform can be represented by a real function (a hologram). However when an asymmetrical mask is used to eliminate the twin image, the symmetry is lost and the result of the inverse Fourier transform is a complex array. The same algorithm can be used to perform the numerical reconstruction, but using this complex array rather than a real array.

This method of spatial filtering has been applied to the hologram of the resolution target that was originally displayed in Fig. 2.8(b) in order to illustrate the removal of the zero-order and twin images. The results are shown in Fig. 3.11.

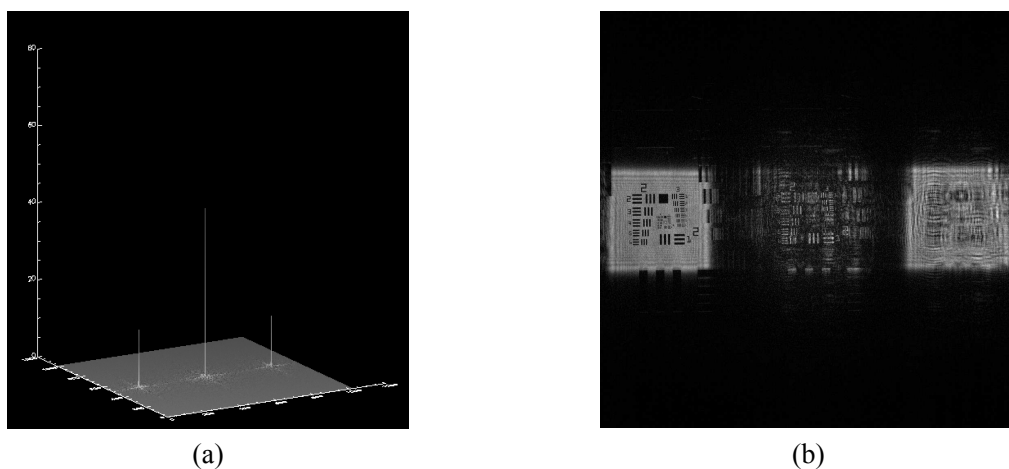
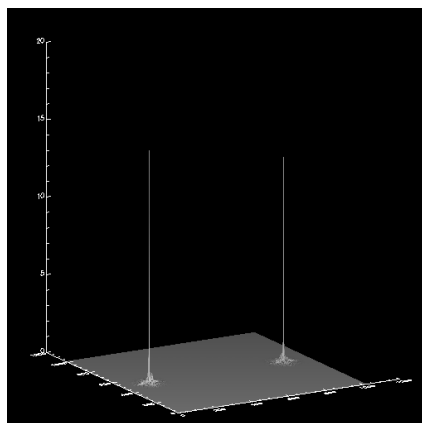
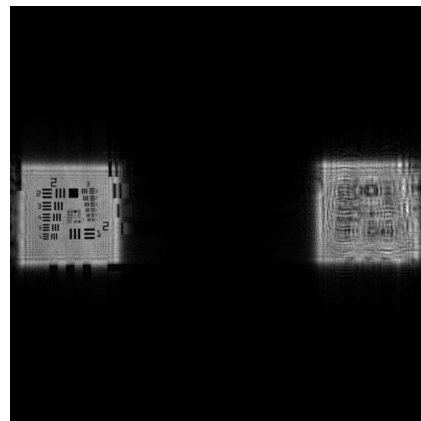


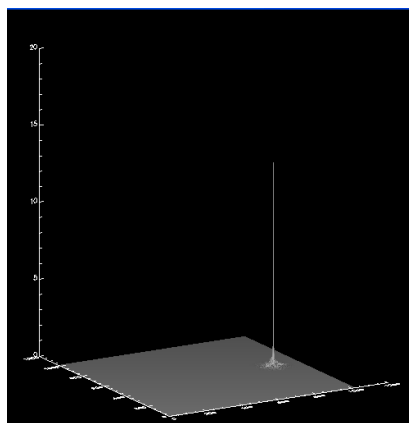
Fig. 3.11 Elimination of the zero-order term and the twin image of a transparent resolution target by spatial filtering: (a) 3D Fourier spectrum of the original hologram shown in Fig. 2.8(b); (b) reconstructed amplitude-contrast image of the original hologram.



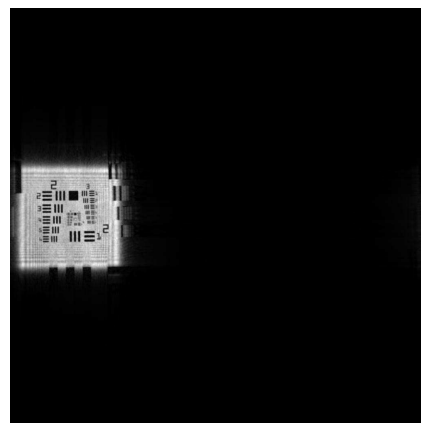
(c)



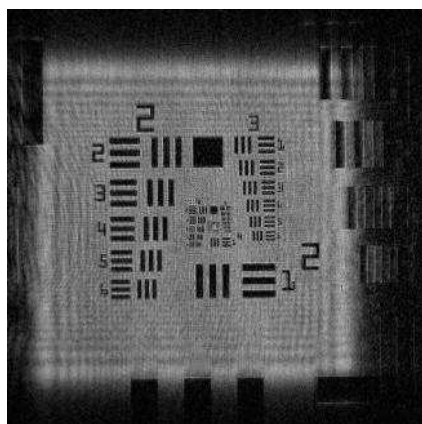
(d)



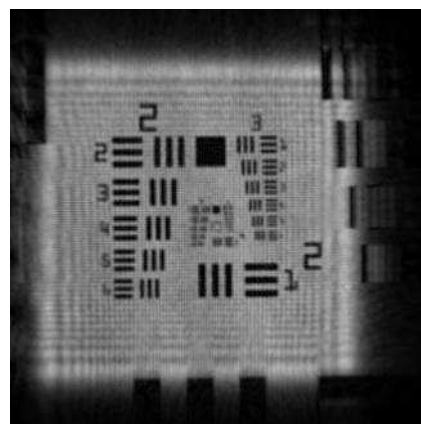
(e)



(f)



(g)



(h)

Fig. 3.11 (Continued) (c) 3D Fourier spectrum of the filtered hologram by a symmetrical mask; (d) reconstructed amplitude-contrast image of the filtered hologram by the symmetrical mask.(e) 3D Fourier spectrum of the filtered hologram by an asymmetrical mask; (f) reconstructed amplitude-contrast image of the filtered hologram by the asymmetrical mask; (g) the enlarged image of resolution target in (b); (h) the enlarged image of resolution target in (d).

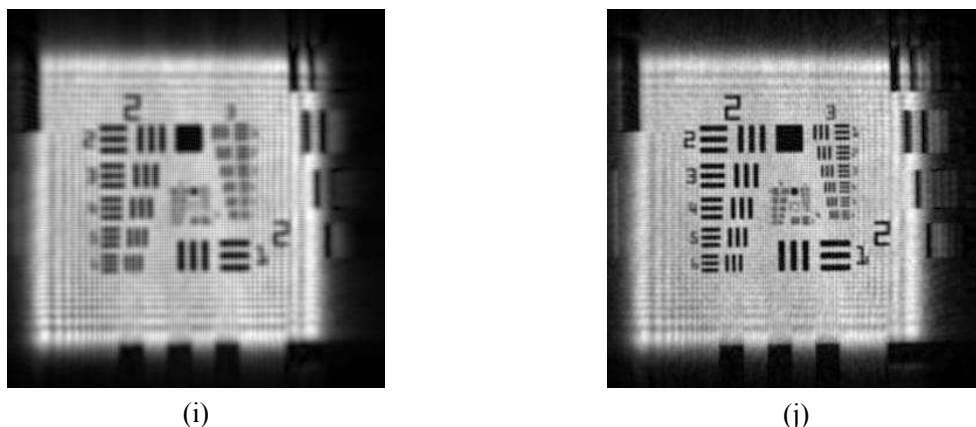


Fig. 3.11 (Continued) (i) the reconstructed result by filtering with an asymmetrical window of 100×100 pixels, only the area of interest is displayed; (j) the reconstructed result by filtering with an asymmetrical window of 200×200 pixels wide, only the area of interest is displayed.

The advantage of the spatial filtering method is apparent in Fig. 3.11. With the removal of the zero-order term and the twin image, the brightness of the image of the resolution target is significantly improved. The background noise has been reduced in Fig. 3.11 (j) so that the image is much cleaner than those found in Fig. 3.11 (g). In this example, a Hanning window has been used to filter the original hologram. The dimensions of the window have to be large enough to avoid the filtering out of the high frequency information from the object. The choice of too small a size of Hanning window will cause the reconstructed image to be blurred. This effect can be clearly seen in Fig. 3.11(i). Meanwhile, by comparing Fig. 3.11(g) and Fig. 3.11(j) carefully, the numbers and the finest bars on the target pattern suggest that some high-frequency information has been lost by the filtering process, even when a large window size of 200×200 pixels is chosen. This method can also be applied to large opaque objects with rough surface, but the loss of high-frequency information is more obvious in the case of well defined test target.

For the case of a large and opaque object, e.g. the die mentioned before, its Fourier spectrum is displayed in Fig. 3.12(a). Besides the central peak of the zero-order term with an amplitude value of about 75, both the ± 1 order terms are very small, with a maximum amplitude value of 0.10 and the frequency information of the object is dispersed over a large spectral range. Hence it is very difficult to know the exact position and shape of the peak of the ± 1 order terms in order to perform effective filtering. This is verified in the results shown in Fig. 3.12. With the highly suppressed noise in the background and the zero-order diffraction, the contrast of the edge of the die in Fig. 3.12(d) and (f) appears too dark in the display because of the filtering. Therefore, in order to preserve the frequency information of the object and remove the unwanted terms, an inverse Hanning window was

applied to suppress the central peak of the zero-order term but retain the other terms intact. The Fourier spectrum of the filtered hologram produced by this method to suppress only the zero-order term is shown in Fig. 3.12(g) and its reconstruction result is shown in Fig. 3.12(h). The corresponding images showing the case of suppression of both the zero-order and the -1 order terms are displayed in Fig. 3.12 (i) and (j), respectively. The results shown in Fig. 3.12 (h) and (j) show better visual effect than those in Fig. 3.12(b), (d) and (f).

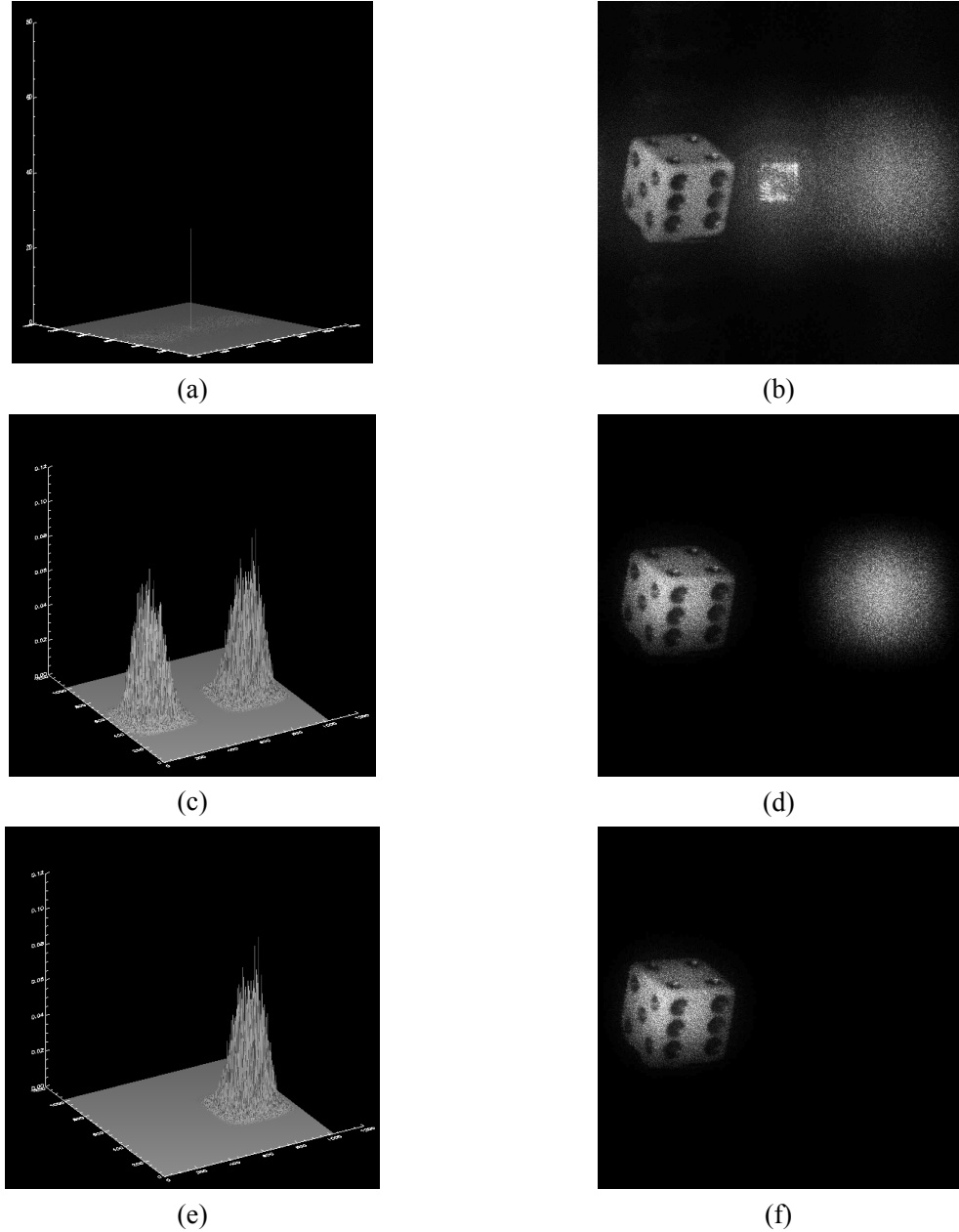


Fig. 3.12 Elimination of the zero-order term and the twin image of a die by spatial filtering: (a) 3D Fourier spectrum of the original hologram shown in Fig. 3.2(a); (b) reconstructed amplitude-contrast image of the original hologram; (c) 3D Fourier spectrum of the filtered hologram by a symmetrical mask; (d) reconstructed amplitude-contrast image of the filtered hologram by the symmetrical mask; (e) 3D Fourier spectrum of the filtered hologram by an asymmetrical mask; (f) reconstructed amplitude-contrast image of the filtered hologram by the asymmetrical mask.

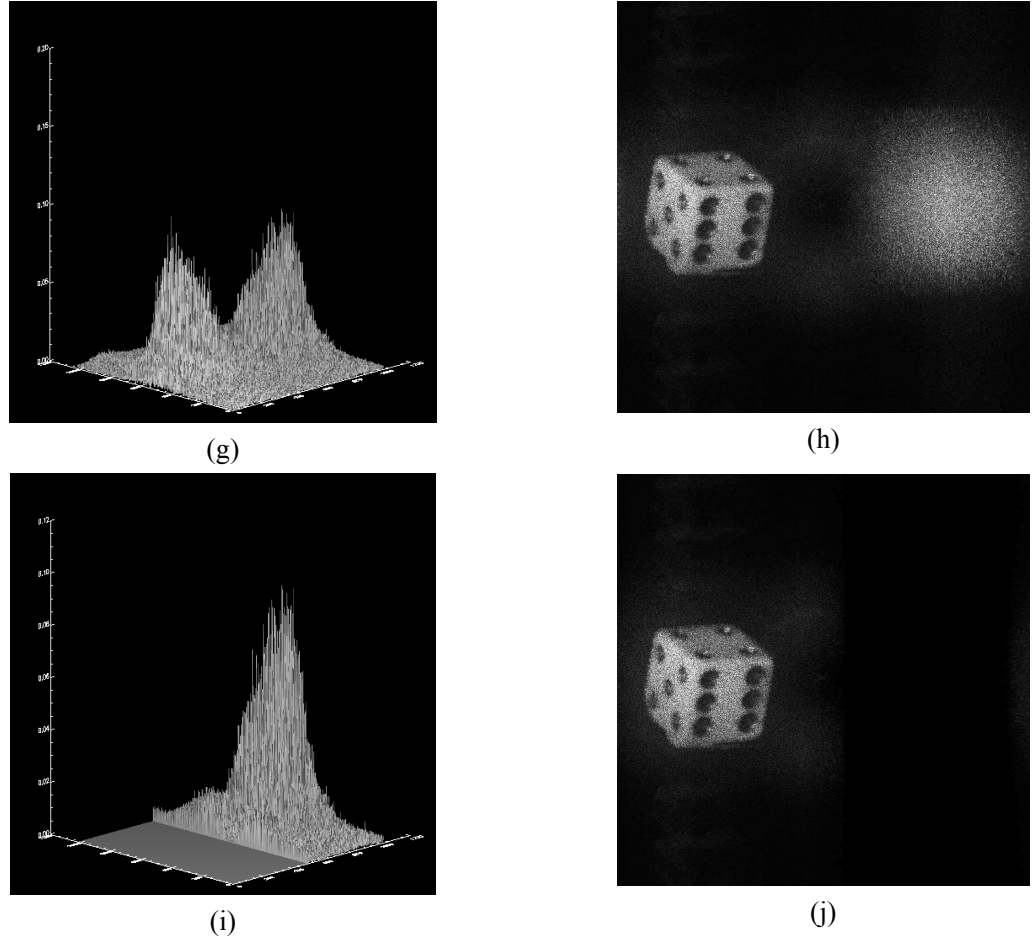


Fig. 3.12 (Continued) (g) 3D Fourier spectrum of the filtered hologram by an inverse Hanning window to suppress zero-order term; (h) reconstructed amplitude-contrast image of the hologram of which Fourier spectrum is shown in (g); (i) 3D Fourier spectrum of the filtered hologram by an inverse Hanning window to suppress zero-order and -1 order term; (j) reconstructed amplitude-contrast image of the hologram of which Fourier spectrum is shown in (i).

3.2.2 Methods based on experiments

The object beam is usually formed from reflected or transmitted light from the object. Due to the coherence of the laser source, a typical object beam contains laser speckle because of the rough nature of the object's surface. In noisy environments, the laser speckle distribution changes stochastically. By subtracting two stochastically changed speckled primary-fringe patterns, it is possible to obtain a clear reconstructed image of the original object. Demoli et al. (2003) presented a fast and simple method to take advantage of this property. The detailed analysis of a quasi-Fourier off-axis setup shows that a much lower average intensity for the subtracted hologram can be expected. If the convolution of the object wave and the quadratic phase factor is small compared to the reference wave, the diffracted intensity is possible to exceed the zero-order in the subtracted hologram. This method requires two holograms to be recorded consecutively under the same conditions,

but with a change in the speckle pattern. If the speckle pattern change is a constant shift in any particular direction, the resulting object reconstruction would not be adequate because of the information loss along the same direction in the output plane. The speckle change must be stochastic in order to suppress the zero-order without information loss.

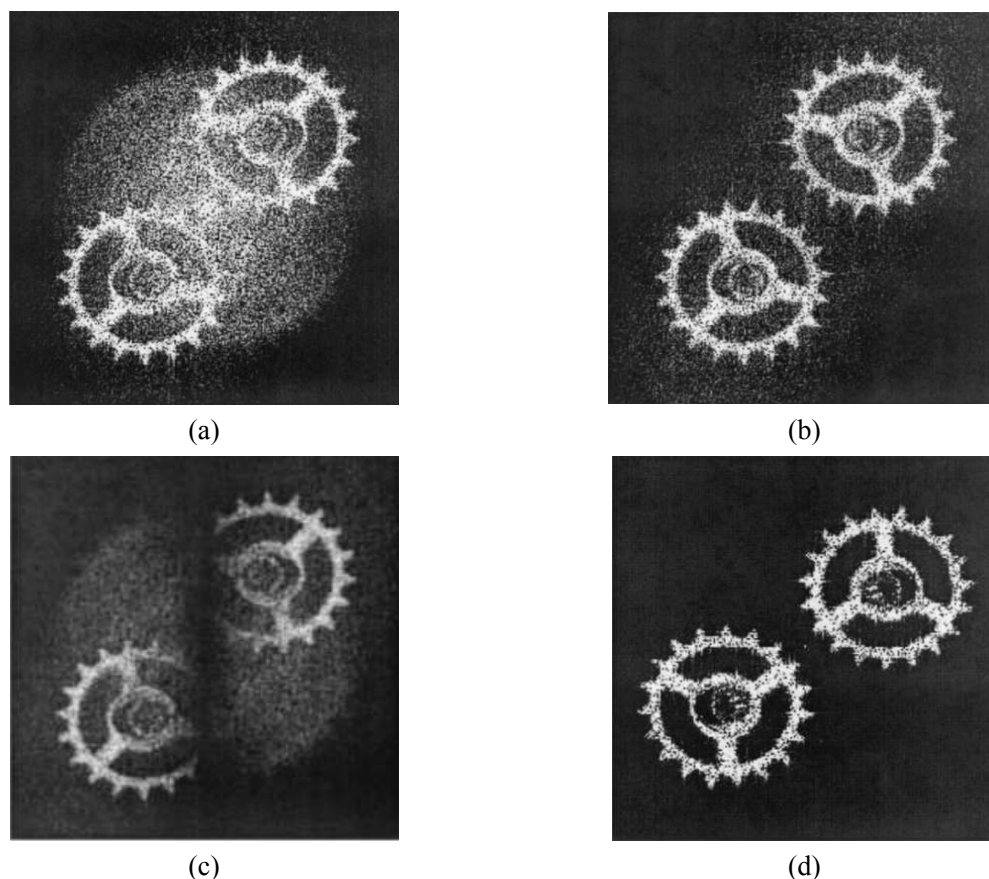


Fig. 3.13 Results of numerical reconstruction: (a) from a single hologram; (b) from a subtracted hologram; (c) from a subtracted hologram for constant shift in the horizontal direction; (d) obtained by subtracting two subsequent holograms from a series. (Reprinted with permission from Demoli et al. (2003). Copyright(2003), Optical Society of America)

Fig. 3.13 shows the results of the subtraction method. Fig. 3.13(a) is the intensity reconstructed image from a single hologram. The zero-order term noticeably affects the reconstructed image. Fig. 3.13(b) is reconstructed from a subtracted hologram. The zero-order term has been greatly suppressed. Fig. 3.13(c) is also reconstructed from a subtracted hologram but there is only a constant shift (1 pixel) between the two holograms in the horizontal direction. This does not improve the quality of the result because of the information loss which is specified as described above. Fig. 3.13(d) is obtained by subtracting two subsequent holograms from a series. To obtain this series of holograms, the vibration mode of the optical table was measured by a digital seismograph. Then a loudspeaker controlled by a computer was used to excite the table in such a way as to

cause a stochastic change of the speckle pattern. The zero-order term is removed and very clear object reconstruction is obtained.

Although this method is very simple and the results in Fig. 3.13 are fairly good, we have not applied it to our experiments for two reasons. The technique requires the vibration properties of the optical bench to be known. A digital seismograph was used in the paper by Demoli et al. (2003). But we do not have the specific equipment to qualify this task.

Takaki et al. proposed some methods based on two shutters and a liquid-crystal phase modulator in order to eliminate both the zero-order and the twin image in hybrid holographic microscopy (Takaki et al., 1999). The two shutters are used to block the object beam and the reference beam independently, and the liquid-crystal phase modulator is used to modulate the phase of the object beam. In this manner the hologram recording parameters are changed. The recording system used is as shown in Fig. 3.14.

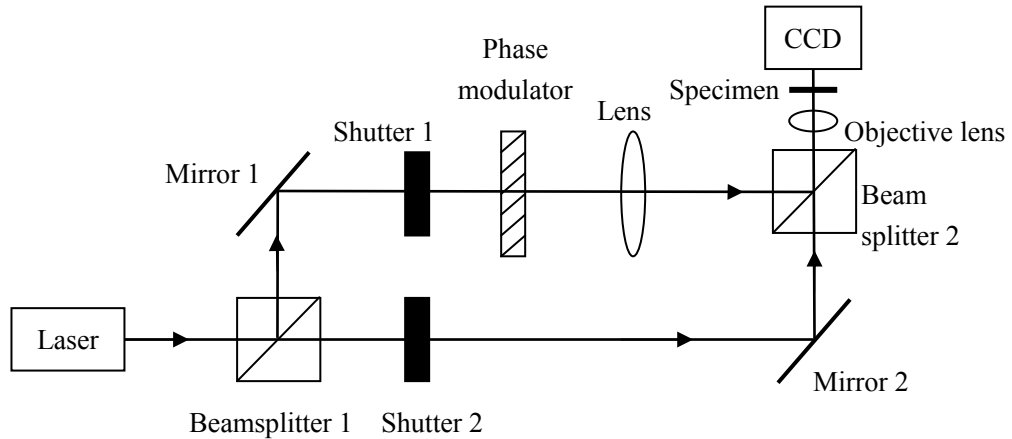


Fig. 3.14 Diagram of the recording system that can eliminate the zero-order term and the twin image

To eliminate the zero-order term, four methods are applied.

Method A: By use of the two shutters, the intensity distribution of both the object beam and the reference beam are obtained and these are subtracted from a hologram to obtain a reconstructed result without the zero-order term. It can be described as

$$U_h - I_o - I_r = U_o U_r^* + U_o^* U_r \quad (3.15)$$

This method requires the capture of three images.

Method B: This method only requires two holograms. Between taking these two holograms, a phase change $\Delta\phi$ is introduced by the liquid-crystal phase modulator. The two holograms can be represented as

$$U_h = I_o + I_r + U_o U_r^* + U_o^* U_r \quad (3.16)$$

$$U_{hp}(\Delta\phi) = I_o + I_r + \exp(i\Delta\phi)U_o U_r^* + \exp(-i\Delta\phi)U_o^* U_r \quad (3.17)$$

Subtracting $U_{hp}(\Delta\phi)$ from U_h , it gives

$$U_h - U_{hp}(\Delta\phi) = [1 - \exp(i\Delta\phi)]U_o U_r^* + [1 - \exp(-i\Delta\phi)]U_o^* U_r \quad (3.18)$$

To use the dynamic range of the image sensor effectively, the magnitude can be maximised when $\Delta\phi = \pi$:

$$[U_h - U_{hp}(\pi)]/2 = U_o U_r^* + U_o^* U_r \quad (3.19)$$

To eliminate both the zero-order term and the twin image, there are also two methods.

Method C: This method adds phase modulation into *Method A* to remove the twin image. It can be written as

$$(U_h - I_o - I_r) - \exp(i\Delta\phi)[U_{hp}(\Delta\phi) - I_o - I_r] = [1 - \exp(i2\Delta\phi)]U_o U_r^* \quad (3.20)$$

The magnitude of the $U_o U_r^*$ term becomes maximum when $\Delta\phi = \pi/2$.

Method D: This method uses only phase modulation. The holograms which are recorded with different phase modulations, ϕ_1 and ϕ_2 , have intensity distributions given by

$$U_{hp}(\phi_1) = I_o + I_r + \exp(i\phi_1)U_o U_r^* + \exp(-i\phi_1)U_o^* U_r \quad (3.21)$$

$$U_{hp}(\phi_2) = I_o + I_r + \exp(i\phi_2)U_o U_r^* + \exp(-i\phi_2)U_o^* U_r \quad (3.22)$$

These two intensity distributions and the intensity distribution with zero phase modulation U_h are used to obtain the $U_o U_r^*$ term:

$$\frac{U_{hp}(\phi_1) - U_h}{\exp(-i\phi_1) - 1} - \frac{U_{hp}(\phi_2) - U_h}{\exp(-i\phi_2) - 1} = \left[\frac{\exp(i\phi_1) - 1}{\exp(-i\phi_1) - 1} - \frac{\exp(i\phi_2) - 1}{\exp(-i\phi_2) - 1} \right] U_o U_r^* \quad (3.23)$$

The magnitude of which becomes maximum when $\varphi_1 = 2\pi/3$ and $\varphi_2 = -2\pi/3$. Three holograms with different phase modulations are required.

Fig. 3.15(a) shows the reconstructed image produced by using *Method A*. Fig. 3.15(b) shows the result reconstructed from the same hologram but this time produced using *Method B*. Fig. 3.15 (b) is less noisy than Fig. 3.15(a). The zero-order term in the centre is less obvious in Fig. 3.15(b). The recorded microscopic object was a type of filamentous green algae, known as *Spirogyra* sp. The spiral structure was observed in both Fig. 3.15(a) and Fig. 3.15(b). Fig. 3.15(c) is the reconstructed image by use of *Method C*. Fig. 3.15(d) is reconstructed from the same hologram by use of *Method D*. A horsetail spore, *equisetum arvense*, was used as the microscopic object. Clearly, the image in Fig. 3.15(d) is sharper and is exempt from the influence of both the zero-order term and the twin image. These images were produced by Takaki et al. (1999).

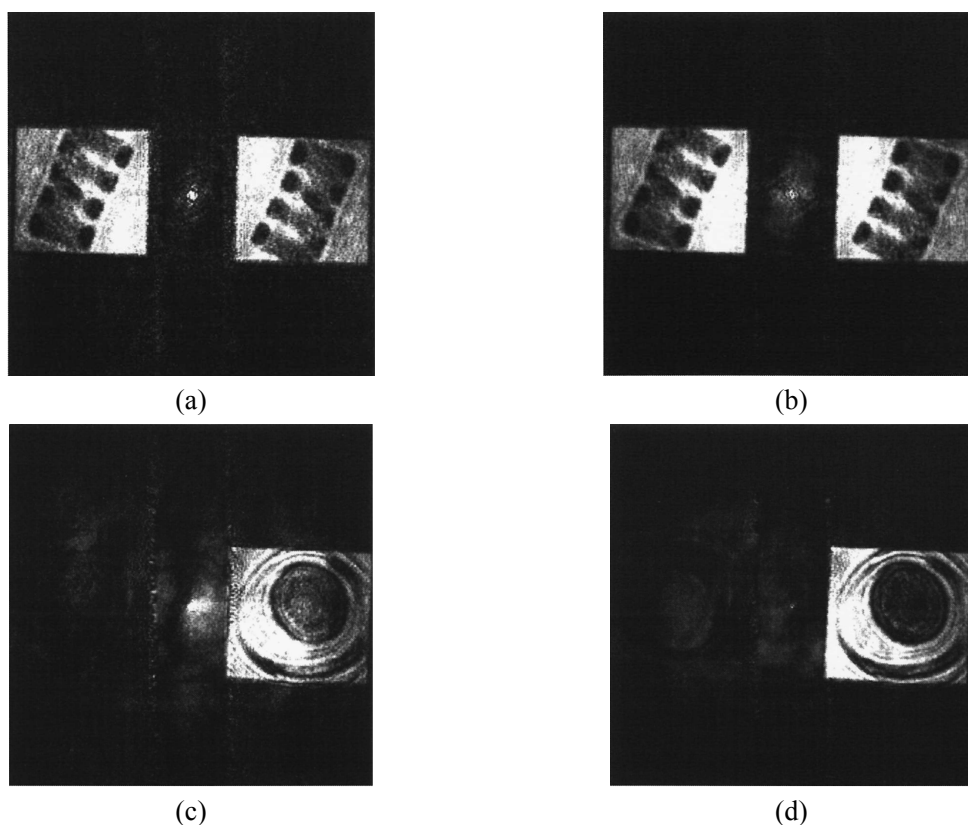


Fig. 3.15 Reconstructed images by use of (a) method A; (b) method B; (c) method C; (d) method D (Reprinted with permission from Takaki et al. (1999). Copyright (1999), Optical Society of America)

Our experimental results using these methods are shown in Fig. 3.16. The results are quite different from the conclusion given by Takaki et al. (1999). The result of using *Method A* in

Fig. 3.16(b) is much worse than it should be according to Eq. (3.15). And it is similar to the result of the die in Fig. 3.16(g). We suppose it results from the weak intensities of the object wave and the reference wave. The background noise of the CCD is quite comparable with the object wave and the reference wave when both of them are rather weak. Therefore the subtraction of the image illuminated only by reference wave and the image illuminated only by object wave causes the extra errors in the reconstructed results. The zero-order term and the twin image were not removed successfully in Fig. 3.16(d) and (e). This also could be the effect of the noise. Moreover, according to the characteristics of the LC phase retarder, shown in Fig. 3.17, the retardance change is very steep in the range 0.2 to 0.8 of a wave of retardance. Any minor deviation of the voltage applied to the LC phase retarder within this range might cause large errors. Therefore it is possible for the error in the phase retardance to cause the incomplete compensation of the phase-shifting holograms, so that both the zero-order term and the twin image have not been suppressed effectively. It is also worth pointing out that the methods were originally applied to microscopy for objects of the scale of microns, rather than the large objects used in our applications with a size of a few millimetres.

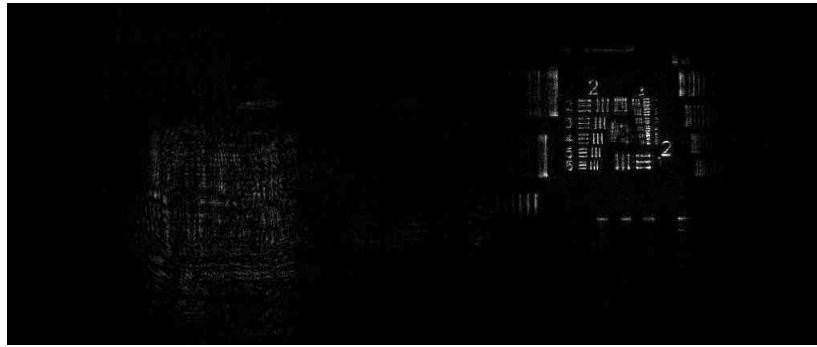


(a)



(b)

Fig. 3.16 The reconstructed intensity-contrast image of a negative resolution test target with zero-order suppression by (a) subtracting the average value of the hologram; (b) *Method A*.



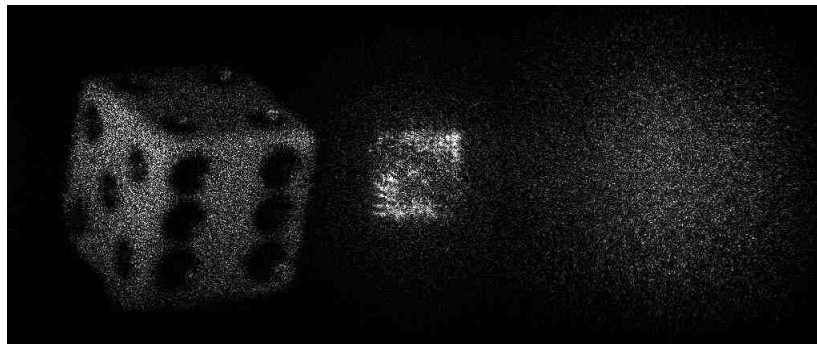
(c)



(d)

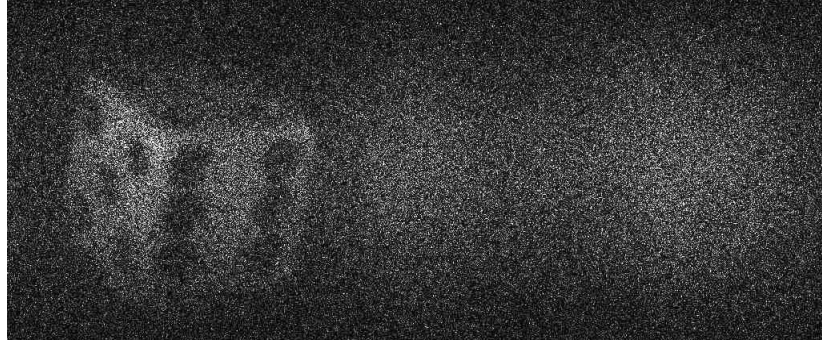


(e)



(f)

Fig. 3.16 (Continued) The reconstructed intensity-contrast image of a negative resolution test target with zero-order suppression by (c) *Method B*; (d) *Method C*; (e) *Method D*; (f) the reconstructed intensity-contrast image of a die with zero-order suppression by subtracting the average value of the hologram.



(g)

Fig. 3.16 (Continued) The reconstructed intensity-contrast image of a die with zero-order suppression by (g) *Method A*.

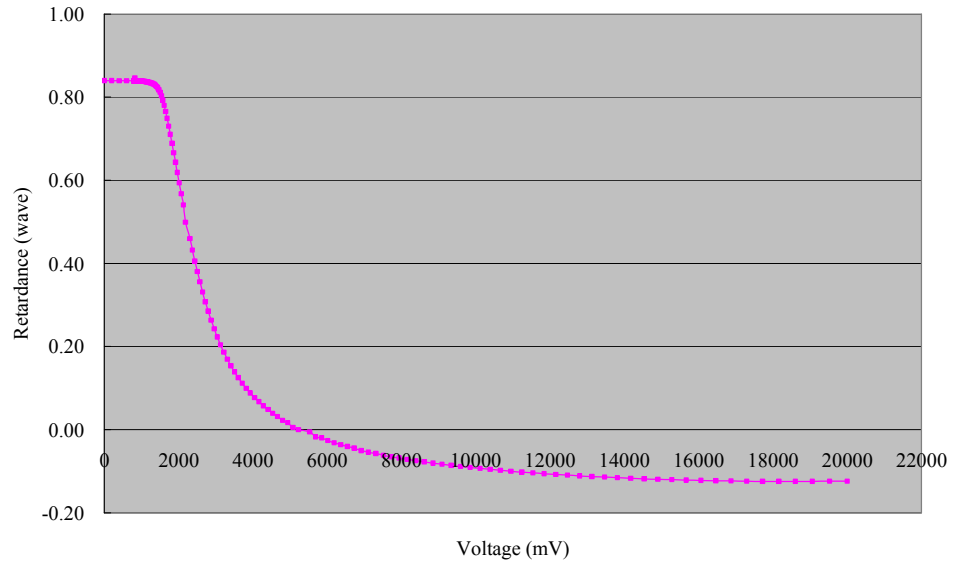


Fig. 3.17 The voltage-to-retardance property of LC variable retarder

The methods that use a phase modulator not only yield reconstructed images that are less noisy than the images produced by the methods that use shutters but they also require fewer images to be captured by the image sensor.

Zhang et al. (2004) used a holographic diffraction grating to act as a beam splitter to construct an off-axis holography system. A diagram illustrating this system is shown in Fig. 3.18.

In this system the holographic grating is illuminated by a plane wave. The zero-order diffraction wave does not change direction and propagates straight through to the CCD sensor as the reference wave. However the +1 order diffraction wave is deflected at a small

angle θ_g and is used to illuminate the object. A mirror is used to direct the object wave to the CCD faceplate to generate the hologram by interference with the reference wave. The principle of this method is similar to *Method B* that was proposed by Takaki et al. (1999) but the redistribution of the object beam on the hologram plane results from a minor adjustment of the reflecting mirror.

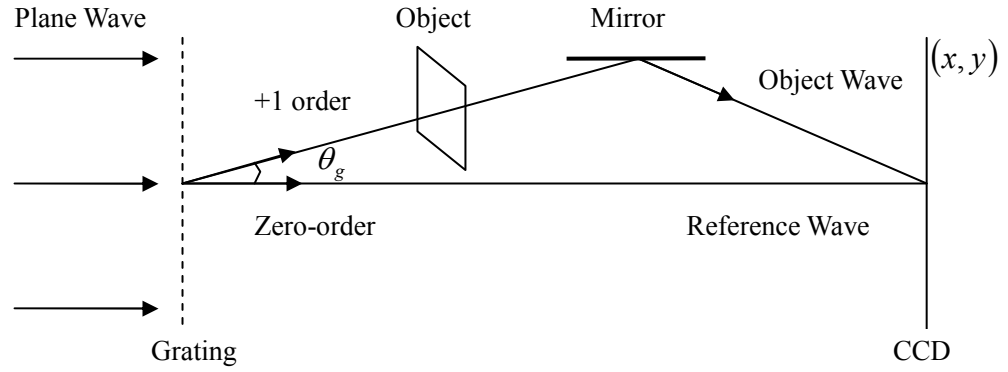


Fig. 3.18 Schematic diagram of the recording system of an off-axis hologram by CCD

Substantially, this method uses a phase-shifting technique. It is easy to obtain phase shifting of the object beam via adjustment of the reflecting mirror. The advantages of this method are that the optical setup and the experimental procedure are simple. However, the method has some disadvantages which are difficult to resolve. Firstly, the detected object should be transparent according to the geometry shown in Fig. 3.18. To secure $|U_{o1}(x, y)| = |U_{o2}(x, y)|$, where $U_{o1}(x, y)$ and $U_{o2}(x, y)$ are the distributions of the object beam on the hologram plane relating to the different positions of the reflecting mirror, a reflective object cannot be used as it will not be able to meet this requirement with the scattered light from object surface. Secondly, this paper did not specify to what extent the reflecting mirror has to be adjusted in order to eliminate the zero-order term. From their preliminary results, reproduced here in Fig. 3.19, the adjustment of the mirror depends on the setup, the position and orientation of the object.

The zero-order term dominated the reconstructed result in Fig. 3.19(a) so that the image of the object was invisible. In Fig. 3.19(b), the zero-order term is suppressed to some extent and therefore the image of the object is revealed. But the zero-order term still remains because $|U_{o1}(x, y)| \neq |U_{o2}(x, y)|$. The visible fringes on the reconstructed images are generated by the interference between the two real images and two conjugate images in the

two holograms used to produce the subtracted hologram. The phase difference $\Delta\phi(x, y)$ relates to the fringes spacing. Fig. 3.19(b) shows that $\Delta\phi(x, y)$ is not small enough so that the fringes can be clearly seen. In Fig. 3.19(c), the zero-order term in the centre has disappeared and both the image of the object and its twin image are displayed clearly without fringes. It shows that in this case $|U_{o1}(x, y)| = |U_{o2}(x, y)|$ and also that the phase difference between the two digital holograms which are being subtracted is constant.

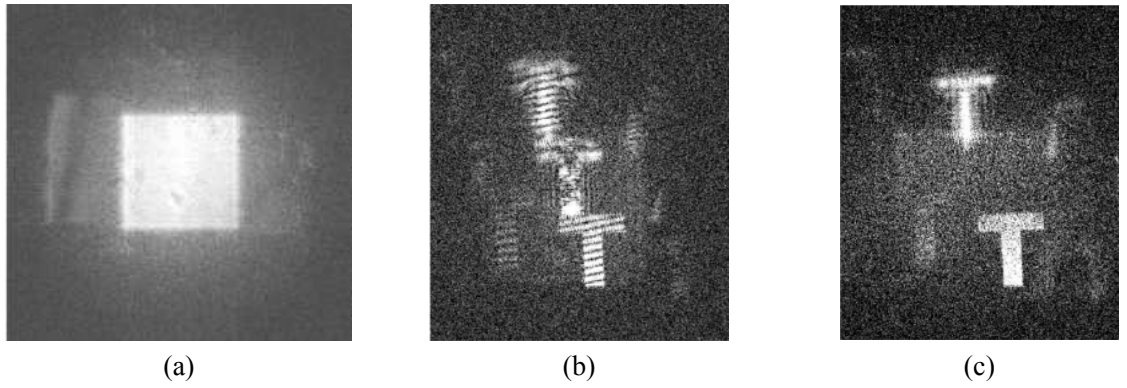


Fig. 3.19 The reconstructed amplitude images: (a) reconstructed from single off-axis hologram; (b) reconstructed from a subtracted hologram without full elimination of the zero-order term; (c) reconstructed from a subtracted hologram with full elimination of the zero-order term (Reprinted from (Zhang et al., 2004). Copyright (2004), with permission from Elsevier).

3.3 Zero-order and twin image suppression for in-line holograms

In-line holography was originally proposed by Gabor in order to improve the resolution of electronic microscopes. Although its inherent twin-image problem was resolved by off-axis holography, it is still widely used in particle-field analysis and soft X-ray holography, due to the simplicity of its recording set-up. However, because of the significantly lower resolution of electronic image sensors when compared to high-resolution photographic film, in-line holography also plays an important role in digital holography in order to take advantage of the full resolution of the image sensor. The use of an in-line hologram necessitates digital filtering in order to suppress the conjugate image and the zero-order image. Wavelength-scanning holography by an argon-ion laser with a tunable intracavity etalon was also proposed (Marron and Schroeder, 1992), but it requires an especially high computation load because of the 3-D Fourier transform including the wavelength axis.

Yamaguchi and Zhang (1997) measured the complex amplitude of the object wave at the CCD plane for an in-line setup by using phase-shifting interferometry. The phase of the reference wave was changed in a stepwise fashion, and the resulting three or four interference fringe images are processed by a computer to yield the distribution of the complex amplitude of the object wave. Then the distribution is Fresnel transformed in the computer to reconstruct images at arbitrary planes. The arrangement is shown in Fig. 3.20.

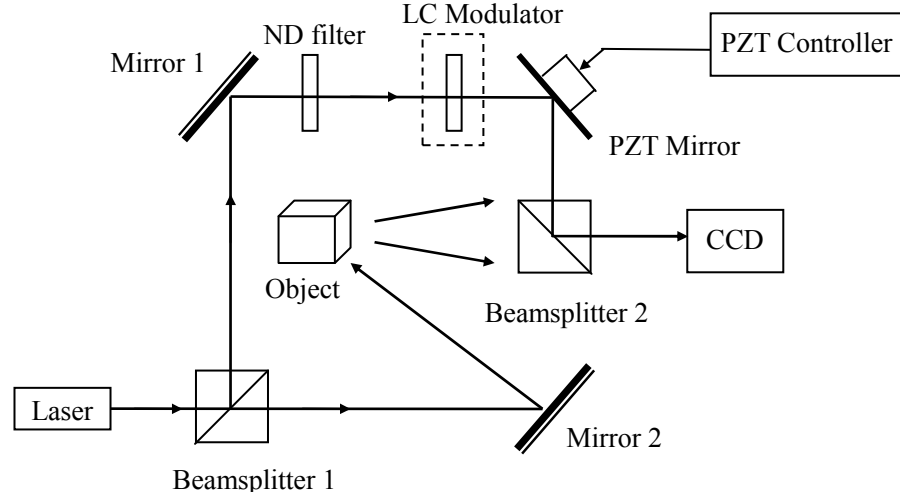


Fig. 3.20 Arrangement for phase-shifting digital holography for measuring diffusely reflecting object

At the CCD plane, the complex amplitudes of the reference wave and the object wave are denoted by $U_r(x, y) = E_r \exp[i\phi_r(x, y)]$ and $U_o(x, y) = E_o(x, y) \exp[i\phi_o(x, y)]$, respectively. The intensity to be recorded is represented by

$$\begin{aligned} U_h(x, y; \delta) &= |U_r(x, y) \exp(i\Delta\phi) + U_o(x, y)|^2 \\ &= |E_r|^2 + |E_o|^2 + U_o U_r^* \exp(-i\Delta\phi) + U_o^* U_r \exp(i\Delta\phi) \end{aligned} \quad (3.24)$$

where $\Delta\phi$ is the phase shift that is provided by the PZT mirror, or alternatively by the LC modulator shown in the frame with dashed lines. The recording of three phase-shifted holograms, obtained using phase-shifts of $\Delta\phi = 0, \pi/2$ and π yields the complex amplitude of the object wave, which is given by

$$U_o(x, y) = \frac{1-i}{4U_r^*} \left\{ U_h(x, y; 0) - U_h\left(x, y; \frac{\pi}{2}\right) + i \left[U_h\left(x, y; \frac{\pi}{2}\right) - U_h(x, y; \pi) \right] \right\} \quad (3.25)$$

For the recording of four phase-shifted holograms using phase-shifts of $\Delta\phi = 0, \pi/2, \pi$ and $3\pi/2$, the complex amplitude of the object wave is given by:

$$U_o(x, y) = \frac{1}{4U_r^*} \left\{ U_h(x, y; 0) - U_h(x, y; \pi) + i \left[U_h\left(x, y; \frac{\pi}{2}\right) - U_h\left(x, y; \frac{3\pi}{2}\right) \right] \right\} \quad (3.26)$$

By carrying out a Fresnel transformation on either Eq. (3.25) or Eq. (3.26), the complex amplitude obtained at a distance Z from the CCD is obtained given by:

$$U_o(\xi', \eta'; Z) = \iint U_o(x, y) \exp \left[ik \frac{(\xi' - x)^2 + (\eta' - y)^2}{2Z} \right] dx dy \quad (3.27)$$

with $k = 2\pi/\lambda$. Setting $Z = -d$ here leads to the complex amplitude at the object plane. For a numerical calculation, Eq. (3.27) can be regarded either as a Fourier transform or as a convolution integral. The former could be implemented using a single fast Fourier transform (FFT), while the latter requires two or three FFTs. The sampling pitch of the reconstructed image is proportional to the wavelength and the object distance in the former case and fixed to the pixel pitch of the CCD in the latter case as illustrated in the previous chapter.

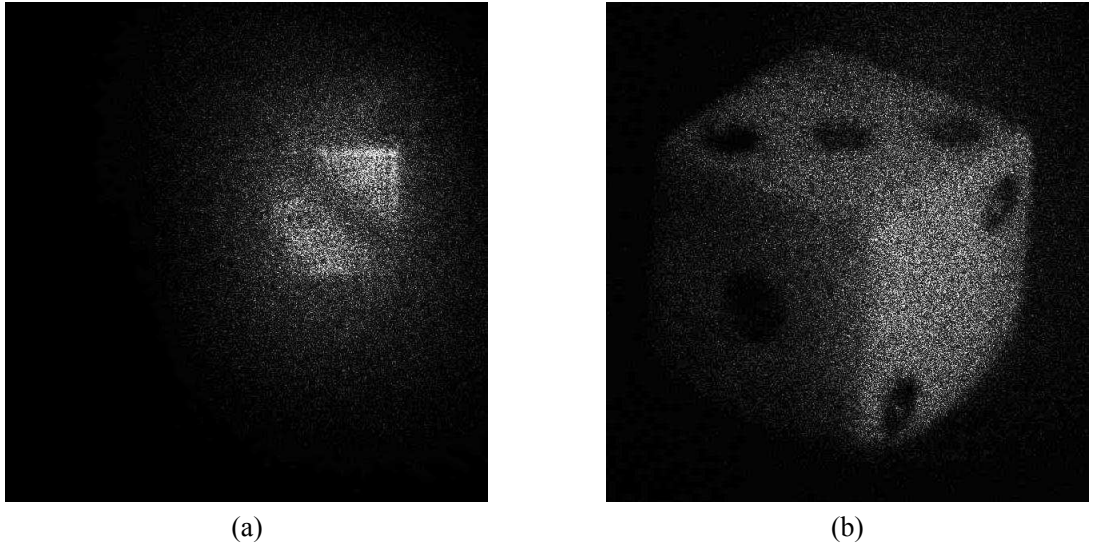


Fig. 3.21 Numerical reconstructed intensity-contrast images of a die by (a) a single in-line hologram, (b) four phase-shifted holograms.

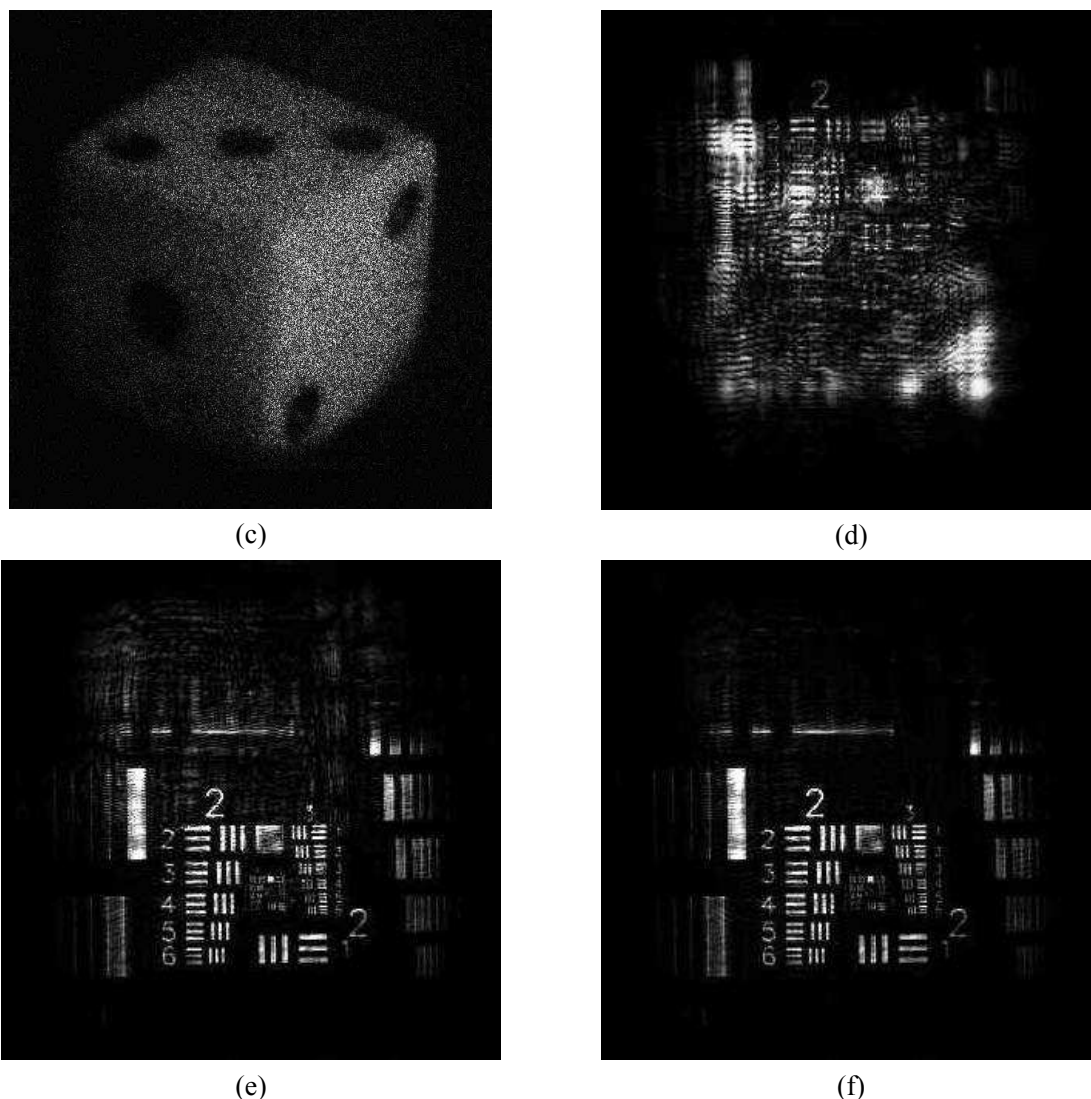


Fig. 3.21 (Continued) (c) Numerical reconstructed intensity-contrast images of a die by three phase-shifted holograms; numerical reconstructed intensity-contrast images of a photographically-negative resolution test target by (d) a single in-line hologram, (e) four phase-shifted holograms and (f) three phase-shifted holograms.

Fig. 3.21(a) to (c) shows reconstructed images of a diffusely reflecting die, produced using an in-line geometry by use of both phase-shifting holography and also conventional digital holography. The die, shown in Fig. 2.6(a), has dimensions of $13\text{mm} \times 13\text{mm} \times 13\text{mm}$ and its centre was positioned at a distance of $d = 415\text{mm}$ from the CCD sensor. Comparing Fig. 3.21 (a) to (c), it can be seen that the bright zero-order image and twin image in Fig. 3.21(a) significantly disrupts the display of the image of the die. However these were both removed by both versions of the phase-shifting method to give a very clear image of the die. Fig. 3.21(d) to (f) show reconstructed images of a photographically-negative resolution test target produced firstly by conventional digital holography and secondly by use of two versions of phase-shifting holography. The resolution target that was shown previously in Fig. 2.8(a) was positioned at a distance of $d = 370\text{mm}$ from the CCD sensor. The

application of the phase-shifting methods greatly reduces the disturbance caused by the zero-order and the twin image, as can be seen in Fig. 3.21(e) and (f).

The results from the version using three phase-shifting holograms in Fig. 3.21 present better suppression of the undesired terms. Theoretically, the results from both the four and the three bucket phase-shifting methods should be the same. However in practice, 4-step phase-shifting operation introduces larger error than 3-step phase-shifting operation. The characteristic curve for the phase retarder shown in Fig. 3.17 does not give the exact voltage to generate an exact phase retardance of 0.25, 0.5 and 0.75 wave. Therefore the applied voltage has to be calculated by interpolation, according to the adjacent known points. Unfortunately the LC phase retarder is a non-linear device, which may cause the interpolated result to deviate from the true value. This is also one reason why the results from different phase-shifting holograms exhibit differences. For real-time applications, phase-shifting digital holography with three holograms is preferred, as it takes less time to capture and process the holograms.

Phase-shifting digital holography realizes 3D imaging for in-line holography and produces high quality images without using imaging lenses. Numerical focusing provides high flexibility in the display of 3D objects. Although the computational load is heavy for the large number of pixels required for high-quality imaging, this difficulty could be overcome by use of a parallel array processor to implement the Fresnel transformation.

Lai et al. (2000) also presented a phase-shifting method to remove the zero-order term and the twin image in an in-line hologram. The system geometry to detect the diffusely reflecting object is similar to that presented in Fig. 3.20. The system arrangement to detect a transparent object is shown in Fig. 3.22.

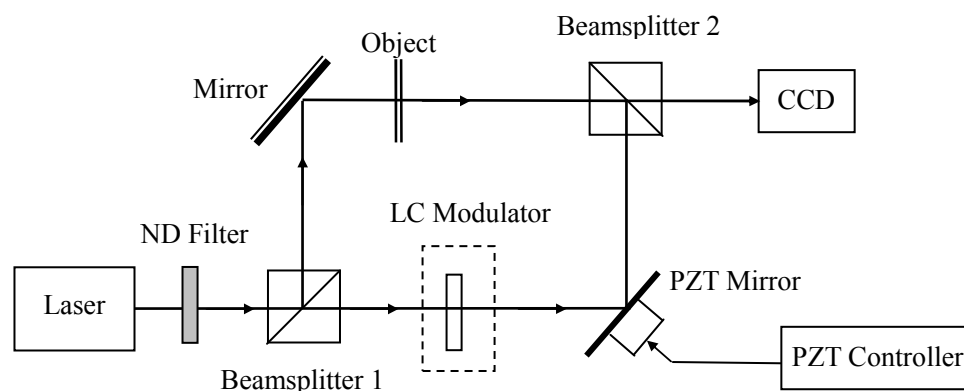


Fig. 3.22 Experimental setup of phase-shifting digital holography for transparent objects

A liquid crystal retarder is used in the reference beam path to vary the phase of the reference wave by $\Delta\varphi = 0, \pi/2, \pi$ and $3\pi/2$. Four quadrature-phase-shifting holograms are achieved. The object wave is denoted as $U_o(x, y) = E_o(x, y)\exp[i\varphi(x, y)]$. Four plane reference plane waves are used, which are

$$\begin{aligned} U_{r1} &= E_r \exp(i0) = E_r, & U_{r2} &= E_r \exp(i\pi/2) = iE_r, \\ U_{r3} &= E_r \exp(i\pi) = -E_r, & U_{r4} &= E_r \exp(i3\pi/2) = -iE_r. \end{aligned}$$

Then the four recorded holograms can be described as

$$U_{h1} = E_r^2 + E_o^2(x, y) + E_r E_o(x, y)\exp[i\varphi(x, y)] + E_r E_o(x, y)\exp[-i\varphi(x, y)] \quad (3.28)$$

$$U_{h2} = E_r^2 + E_o^2(x, y) - iE_r E_o(x, y)\exp[i\varphi(x, y)] + iE_r E_o(x, y)\exp[-i\varphi(x, y)] \quad (3.29)$$

$$U_{h3} = E_r^2 + E_o^2(x, y) - E_r E_o(x, y)\exp[i\varphi(x, y)] - E_r E_o(x, y)\exp[-i\varphi(x, y)] \quad (3.30)$$

$$U_{h4} = E_r^2 + E_o^2(x, y) + iE_r E_o(x, y)\exp[i\varphi(x, y)] - iE_r E_o(x, y)\exp[-i\varphi(x, y)] \quad (3.31)$$

Illuminating the four holograms described above with the corresponding reference waves, we have

$$U_{r1}U_{h1} = E_r^3 + E_r E_o^2(x, y) + E_r^2 E_o(x, y)\exp[i\varphi(x, y)] + E_r^2 E_o(x, y)\exp[-i\varphi(x, y)] \quad (3.32)$$

$$U_{r2}U_{h2} = iE_r^3 + iE_r E_o^2(x, y) + E_r^2 E_o(x, y)\exp[i\varphi(x, y)] - E_r^2 E_o(x, y)\exp[-i\varphi(x, y)] \quad (3.33)$$

$$U_{r3}U_{h3} = -E_r^3 - E_r E_o^2(x, y) + E_r^2 E_o(x, y)\exp[i\varphi(x, y)] + E_r^2 E_o(x, y)\exp[-i\varphi(x, y)] \quad (3.34)$$

$$U_{r4}U_{h4} = -iE_r^3 - iE_r E_o^2(x, y) + E_r^2 E_o(x, y)\exp[i\varphi(x, y)] - E_r^2 E_o(x, y)\exp[-i\varphi(x, y)] \quad (3.35)$$

Adding Eqs. (3.32) - (3.35) together, we obtain

$$U_{r1}U_{h1} + U_{r2}U_{h2} + U_{r3}U_{h3} + U_{r4}U_{h4} = 4E_r^2 E_o(x, y)\exp[i\varphi(x, y)] \quad (3.36)$$

In Eq. (3.36), it is clear that only the pure wavefront of the object detected on the CCD faceplate remains and the other unwanted terms are cancelled out. If then the Fresnel approximation or convolution approach is applied to Eq. (3.36), the complex amplitude of the object wave at any distance from the CCD faceplate can be reconstructed.

This method is simple in principle, but very hard to realise practically for conventional holography. However via its digital recording on the CCD sensor and numerical reconstruction performed in the computer, it can be performed very easily by digital holography. There are at least two errors in this method which degrade the quality of the

reconstructed image. One is the noise from the CCD sensor and the second noise source is from the phase-shift error introduced by the phase-shifting devices. Both of these are assumed to have a normal distribution with a mean value of zero and a standard deviation of σ for the former and $\Delta\varphi$ for the latter. So they can be reduced by averaging multiple holograms for each phase step.

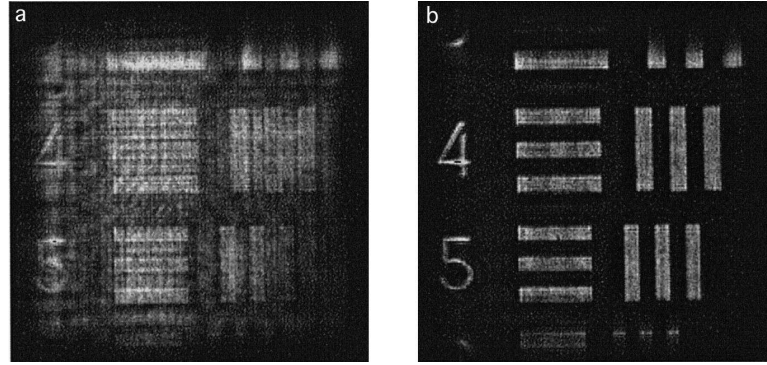


Fig. 3.23 Reconstructed images of a part of a resolution target produced by: (a) single hologram; and (b) four quadrature-phase-shift holograms. (Reprinted from (Lai et al., 2000). Copyright (2000), with permission from Elsevier.)

Fig. 3.23 shows the reconstructed images of a part of a resolution target from both a single hologram and also from four quadrature-phase-shift holograms. By use of the proposed phase-shifting method, the twin image is eliminated and never contaminates the real image.

The same quadrature-phase-shifting holograms that were used to produce the results presented in Fig. 3.21(b) and (e) were now used to perform the phase-shifting method described in Eq. (3.36). Fig. 3.24(a) and (b) show the corresponding results. Through a comparison of Fig. 3.21 and Fig. 3.24, for both the transparent target and the diffusely reflecting die, Yamaguchi and Zhang's method displayed better performance in suppressing the undesired terms, than that of Lai et al.'s phase retrieval approach. Taking a closer look at Eqs. (3.26) and (3.36), it is not difficult to find the reason for this. As described above, the phase error caused by the LC variable retarder, will affect the reconstruction results when using the quadrature-phase-shifting holograms. In Eq. (3.36), the phase-shifted reference waves are simulated by computer employing an ideal formula. However, the reference waves used in the experiment are not ideal. For example, the intensity of the reference wave is not uniform on the CCD sensor, because of dust or blemishes in the beam expander. Also the phase steps of the reference waves are not strictly $\pi/2$. The multiplications and additions of these ideal reference waves and the phase-shifting holograms with errors will surely increase noise and degrade the reconstruction results. This is in accordance with the

results shown in Fig. 3.24. Therefore in the practical application of phase-shifting digital holography, Yamaguchi and Zhang's method is used by the majority of the researchers in this area.

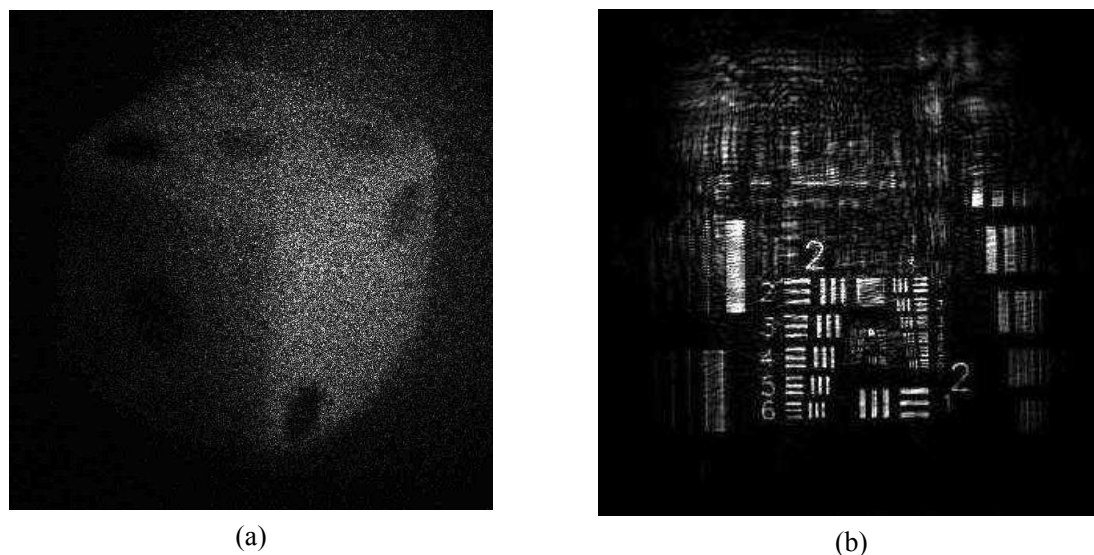


Fig. 3.24 The reconstructed intensity-contrast image by phase-shifting method proposed by Lai et al. (2000): (a) the die reconstructed from the same hologram sequence which were also used to obtain Fig. 3.21(b); (b) the resolution target reconstructed from the same hologram sequence which were also used to obtain Fig. 3.21(e).

3.4 Conclusions

In this chapter, various methods that have been used to suppress the zero-order and twin image terms have been briefly reviewed. These terms severely degrade the results of the in-line holograms. Even in off-axis digital holography, though the different terms are separated by the spatial carrier that has been introduced, the existence of these terms decreases the dynamic range of the reconstruction results. In conventional holography, this problem is mainly resolved by experimental methods. But with the facilities of digital recording and numerical reconstruction, digital holography is more flexible in its capability to suppress the effects of these terms. Digital filtering can easily be integrated into the numerical reconstruction stage to enhance the quality of results, as seen in section 3.2.1. Also with digital methods, the system geometry is simplified. For example, Gabor's two-quadrature-hologram method is the same as the result of adding Eq. (3.32) and (3.33) together. Even so, the configuration of Gabor's interference microscope was very complicated. By comparison the digital recording of four quadrature holograms, the implementation of Eq. (3.36) is straightforward and easy.

Besides digital image processing methods, the phase-shifting method is also widely used in digital holography in order to suppress the unwanted terms. In sections 3.2.2 and 3.3, a phase shift is introduced either in the object beam or the reference beam, depending on the conditions and requirements of the system setup. With either a known, or an unknown phase shift, some undesired terms are cancelled out, which definitely improves the reconstruction results. From the results shown in Fig. 3.21 and Fig. 3.23, it can be seen that phase-shifting digital holography provides a fundamental improvement over conventional digital holography. With this approach, the full resolution of the CCD sensor can be exploited and the quality of the reconstructed image is significantly enhanced. However, these benefits are achieved at the cost of an increased complexity in the whole system. In some real-time applications with high-speed requirements, the necessity of capturing more than three holograms for each state of the object would be a bottle-neck to the use of this technology.

Digital filtering of the digital hologram with an inverse Hanning window is proposed in order to suppress the zero-order term and the twin image described in section 3.2.1. The feasibility and the influences of the various different techniques to suppress the zero-order term and the twin image for the application of macroscopic objects are specified in this chapter.

In our main application of optical contouring by digital holography, which will be discussed later, phase-shifting digital holography was not used. This is because the characteristic curve of the LC variable phase retarder shown in Fig. 3.17, only works for linearly polarised light, whose polarisation direction is 45° to the fast axis of the retarder. However in optical contouring applications, one of the laser systems is designed to couple with an optical fibre. The state of polarisation of the laser beam delivered by the optical fibre is changed to an arbitrary status, dictated by fibre bending and environmental parameters, for example temperature and humidity. Hence, with this set-up phase-shifting using the LC variable retarder cannot be performed. Therefore, off-axis digital holography was used in the optical contouring section described in Chapter 5.

References:

- CHEN, G.-L., LIN, C.-Y., KUO, M.-K. & CHANG, C.-C. (2007) Numerical suppression of zero-order image in digital holography. *Optics Express*, 15, 8851-8856.
- CUCHE, E., MARQUET, P. & DEPEURSINGE, C. (2000) Spatial filtering for zero-order and twin-image elimination in digital off-axis holography. *Applied Optics*, 39, 4070-4075.
- DEMOLI, N., MESTROVIC, J. & SOVIC, I. (2003) Subtraction digital holography. *Applied Optics*, 42, 798-804.
- KREIS, T. M. & JUPTNER, W. P. O. (1997) Suppression of the dc term in digital holography. *Optical Engineering*, 36, 2357-60.
- LAI, S., KING, B. & NEIFELD, M. A. (2000) Wave front reconstruction by means of phase-shifting digital in-line holography. *Optics Communications*, 173, 155-160.
- LEITH, E. N. & UPATNIEKS, J. (1962) Reconstructed wavefronts and communication theory. *Journal of the Optical Society of America*, 52, 1123-1130.
- LIU, C., LI, Y., CHENG, X., LIU, Z., BO, F. & ZHU, J. (2002) Elimination of zero-order diffraction in digital holography. *Optical Engineering*, 41, 2434-2437.
- MARRON, J. C. & SCHROEDER, K. S. (1992) Three-dimensional lensless imaging using laser frequency diversity. *Applied Optics*, 31, 255-62.
- TAKAKI, Y., KAWAI, H. & OHZU, H. (1999) Hybrid holographic microscopy free of conjugate and zero-order images. *Applied Optics*, 38, 4990-6.
- YAMAGUCHI, I. & ZHANG, T. (1997) Phase-shifting digital holography. *Optics Letters*, 22, 1268-1270.
- ZHANG, Y., LU, Q. & GE, B. (2004) Elimination of zero-order diffraction in digital off-axis holography. *Optics Communications*, 240, 261-7.

4 Enhancements of Digital Holography

4.1 Introduction

As the digital version of optical holography, digital holography has inherited both advantages and disadvantages from its optical counterpart. It has the ability to preserve both the intensity and phase information of the wave field, and can be performed within seconds or even in less than a second, without resorting to inconvenient wet chemical processes. However digital holography has also inherited some fundamental problems, such as the presence of the zero-order diffraction and speckle noise. Moreover, due to the characteristics of the recording media in digital holography, the effects of these two problems are more troublesome than in optical holography. The adoption of a CCD sensor as the recording medium facilitates fast and efficient recording and processing of the hologram, however the limited number of pixels in the CCD chip, i.e. the relatively low resolution of the recording media, is the biggest obstacle to the further development of digital holography.

In this chapter, factors that enhance the results of digital holography will be dealt with. Firstly, the influence of the recording conditions will be discussed. Then speckle noise and the methods used to suppress it will be presented. Finally, a method is proposed to overcome the limitations due to the relatively low CCD resolution compared to holographic film.

4.2 Recording conditions

4.2.1 Typical configurations in digital holographic systems

Fig. 4.1 shows some typical recording configurations used in digital holography. In Fig. 4.1(a) and (b), a plane wave has been used as the reference wave. The benefit of using a plane reference wave is that it gives the ability to perform numerical refocusing by computer as part of the numerical reconstruction according to Eq. (2.20). A beamsplitter is used to couple the plane reference wave into the system in Fig. 4.1(b). This allows the object to be positioned symmetrically and easily. Usually there is a stand or equivalent

equipment that is used to hold the object securely. As explained in section 1.5, normally the angle between the object wave and the reference wave is smaller than 4° , depending on the specification of the CCD for off-axis digital holography. If the object is positioned too near the CCD target, the object holder might block the reference wave in a configuration such as that shown in Fig. 4.1(a). Although the use of a beamsplitter introduces some longitudinal image shift, in the application of imaging macroscopic objects, this shift does not have a noticeable effect on the results. Therefore, the configuration shown in Fig. 4.1(b) is frequently chosen by researchers. In our experiments, this configuration has also been used.

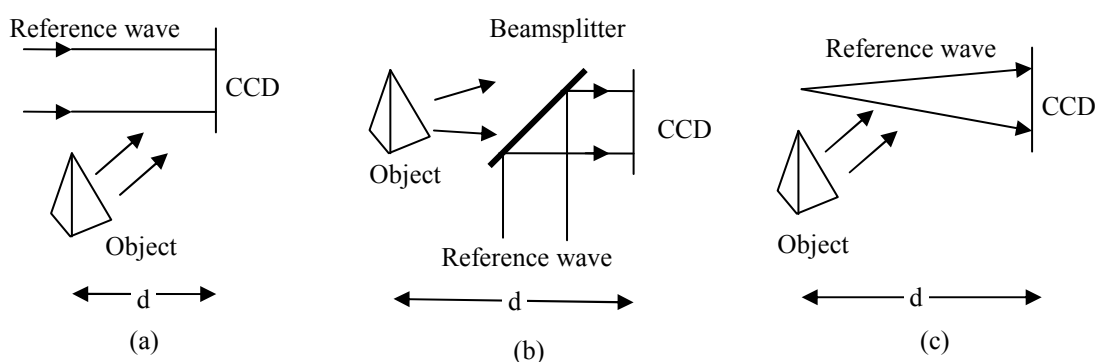


Fig. 4.1 Typical recording configurations in digital holography

Contrary to the situations shown in Fig. 4.1(a) and (b), Fig. 4.1(c) shows the case where a spherical wave has been used as the reference wave in lensless Fourier holography. The principle and example of lensless holography was previously described in section 2.2.4. The angle between the object wave and the reference wave is approximately constant over the full sensor area (see Fig. 4.2(b)). Thus the spatial-frequency spectrum is fully used on every point of the CCD, so that in this case the object can be positioned closer to the CCD sensor. However for a plane reference wave, the angle between the reference beam and the object beam varies over the surface of the sensor (see Fig. 4.2(a)). Therefore in some places the spatial bandwidth is not used fully, assuming that the sampling theorem is considered.

In our experiments, lensless Fourier holography was not used, even though it provides better usage of the spatial-frequency spectrum and involves simpler calculations. This was mainly because it is not possible to perform numerical refocusing using this method which has been analysed in section 2.2.4.

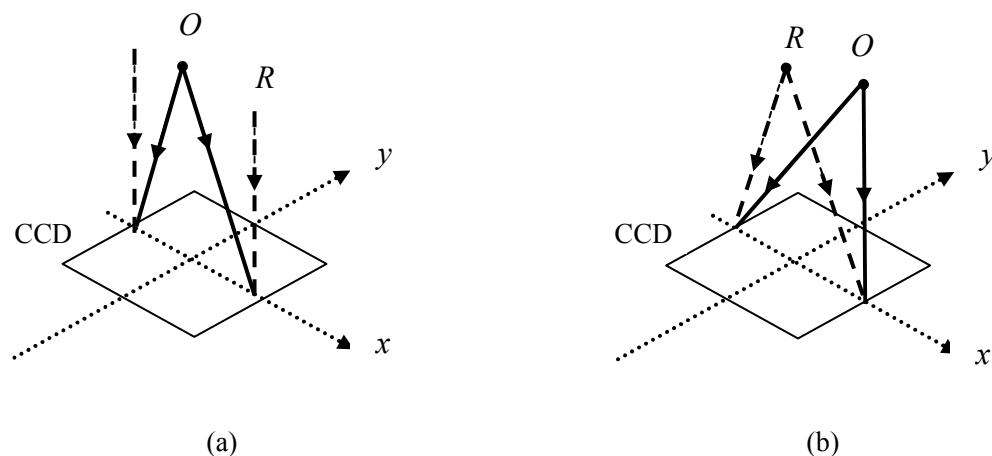
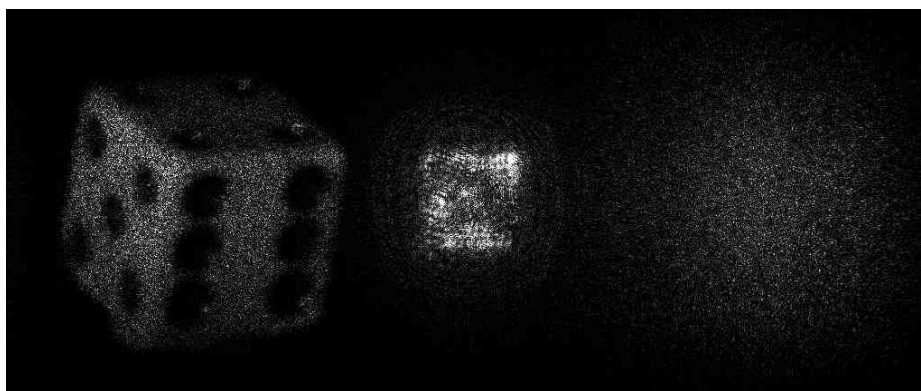


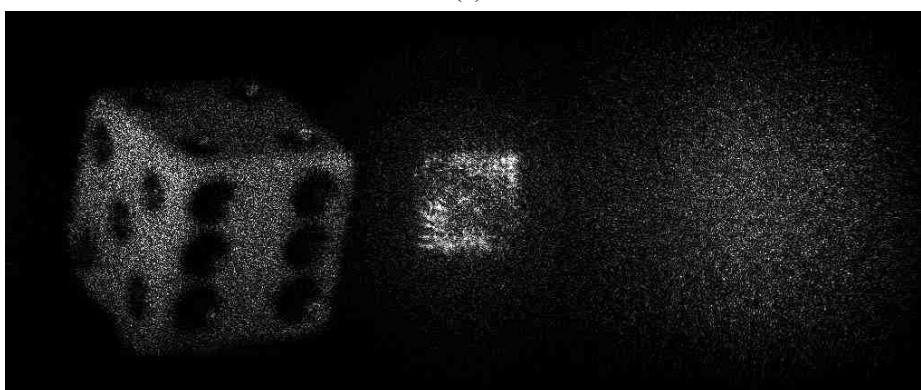
Fig. 4.2 Diagrams to show the angle between the reference wave (R) and the object wave (O) over the surface of the sensor: (a) with plane reference wave; (b) with spherical reference wave.

4.2.2 The intensity of the reference wave

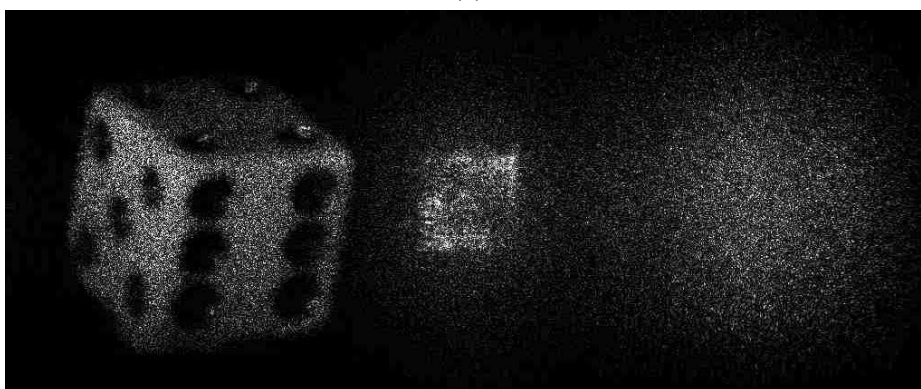
A hologram is the interference pattern formed by superposition of the reference wave and the object wave. According to interference theory, the maximum contrast of an interference pattern is achieved when the intensity of both waves is equal. The influence of the intensity ratio of the reference wave and the object wave is clearly shown in Fig. 4.3, which presents the reconstruction results from different digital holograms, where the intensity ratio has been changed by using different neutral density filters (NDF) to change the intensity of the reference wave. Because the object wave is the light reflected from the surface of the object, in this case a die, which is much weaker than the wave that was used to illuminate it, the NDF used in the optical path of the reference wave has to have a large attenuation factor to make both waves comparable. As well as the different holograms with varying ratios of reference wave and object wave, for each of these intensity ratios, an image of only the reference wave and an image with only the object wave were recorded in order to find out the optimization conditions to obtain good reconstruction results. Table 4.1 gives the results of this investigation.



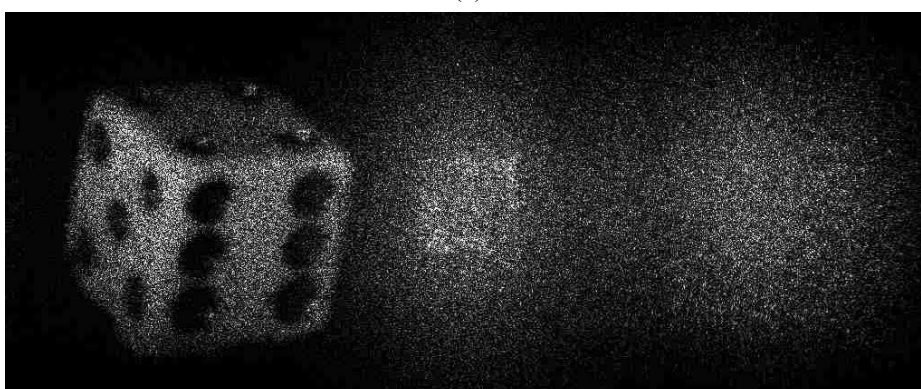
(a)



(b)

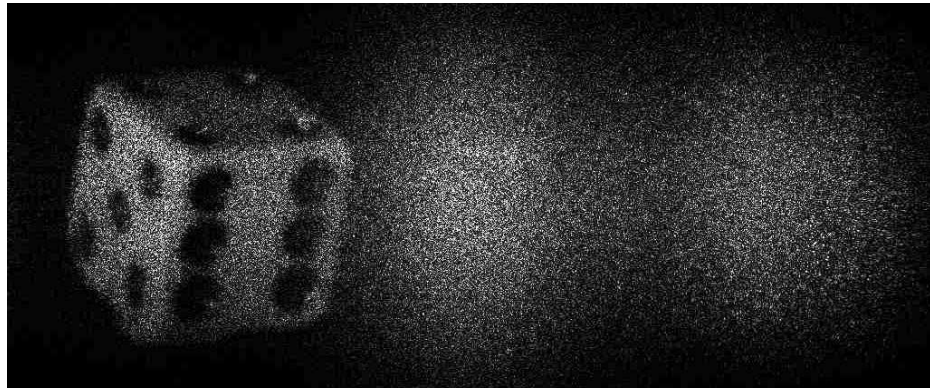


(c)



(d)

Fig. 4.3 The reconstructed images of the die, produced by recording with different intensity ratios between reference wave and object wave. The intensity ratio was changed by using different NDFs in the reference beam path: (a) with a NDF of 0.1%; (b) with two NDFs of 0.1% and 50%; (c) with two NDFs of 0.1% and 33%; (d) with three NDFs of 0.1%, 33% and 50%.



(e)

Fig. 4.3 (Continued) (e) the reconstruction result of the die with two NDFs of 0.1% and 10% in the reference beam path.

	with a NDF of 0.1%	with 2 NDFs of 0.1% and 50%	with 2 NDFs of 0.1% and 33%	with 3 NDFs of 0.1%, 33% and 50%	with 2 NDFs of 0.1% and 10%
Attenuation factor (%)	0.1	0.05	0.033	0.017	0.01
Mean value of the image of the reference wave	95.2464	65.5980	54.4046	38.8397	34.0447
Mean value of the image of the object wave	39.2769	38.9835	39.3064	39.0676	39.2637
Intensity ratio of the reference wave and object wave	2.42500	1.68271	1.38411	0.994166	0.867078

Table 4.1 The mean values and intensity ratios of the reference wave and object wave by using different NDF combinations

It is clear that with greater attenuation of the reference wave, the image of the die is improved and more object detail is revealed. But further attenuation of the reference wave results in the significantly noisy background in the reconstruction results as shown in Fig. 4.3(e). If the reference wave is too intense, the image of the die is too dark and not many details are revealed, as shown in Fig. 4.3(a). Under the dual considerations of both displaying more information from the object and also restricting the background noise, there must be an optimisation of the intensity ratio of the object and reference waves. Comparing the five pictures in Fig. 4.3, this optimisation can be seen from Fig. 4.3(c) and (d). By referring to Table 4.1, it matches with the assumption that the best contrast of the fringes is achieved when the intensities of both the reference wave and the object wave are close to each other. These results show that the intensity ratio of both waves is important in

obtaining a good hologram, so the intensity ratio must be well-controlled in the recording procedure.

Schnars and Juptner (2005) proposed a method to adjust the intensity ratio to the optimised level. An aperture is used to cover one half of the expanded reference wave. The brightness (measured in greyscale value) in that half of the CCD that is being illuminated by the reference and object waves together should be twice as high as the brightness in the other half, which is illuminated only by the object wave. However this is not valid for the conditions in our experiments. The hologram shown in Fig. 4.4(a) is generated by a reference wave and an object wave with an intensity ratio of 1:1. This is measured by blocking either wave to obtain two images with solely the reference wave, or alternatively solely the object wave and then performing calculations in order to get the results similar to those presented in Table 4.1. Then a paper screen was used to block off half of the reference wave and an image is captured as shown in Fig. 4.4(b). The intensity ratio of the right half of Fig. 4.4(b) compared to its left half is 1.4, which is much less than the ratio of 2.0 that was proposed by Schnars and Juptner (2005). This difference might result from the different output powers of the laser sources used. According to Schnars et al. (1996) and other published papers by this group, laser sources with typical output powers of a few hundred milliwatts were used in their systems. However our most powerful laser has an output power of about 40 mw. Therefore, the object wave scattered from the surface of the object is very weak, as shown in the left half of Fig. 4.4(b). For such a very weak signal, the noise from the CCD has a larger effect than it would in the case of a stronger signal. Hence the reference wave and the object wave are preferred to be measured individually to find the optimisation ratio, rather than by using the method provided by Schnars and Juptner (2005).

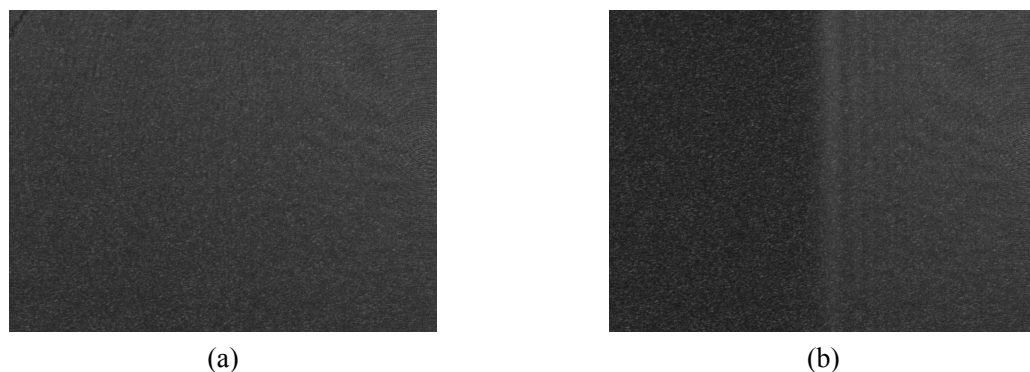


Fig. 4.4 Example to test the intensity ratio of the reference wave and object wave: (a) a hologram with equal intensity of the reference wave and object wave; (b) the image captured by blocking a half of the reference wave.

4.3 Reduction of speckle noise

The experiments of Kozma and Christensen (1976) indicated that the image resolution of a coherently illuminated system is less than half of that produced by an equivalent incoherently illuminated system, when the object is a black and white diffuse grating. This factor falls to less than one-fifth when the object is a continuous grayscale-image. Because speckle noise severely degrades the results of digital holography as seen in previous chapters, speckle noise is an important issue that digital holography must deal with. It essentially arises from the limited size of the CCD camera and the rough object's surface that is being illuminated by the coherent light source. There are many methods that have been used to resolve this problem. Usually these methods can be divided into two categories, digital image processing techniques and optical techniques.

4.3.1 Digital image processing

Garcia-Sucerquia et al. (2005) proposed a digital image processing method to effectively reduce speckle noise that combined a mixed approach of matrix reduction and median filtering.

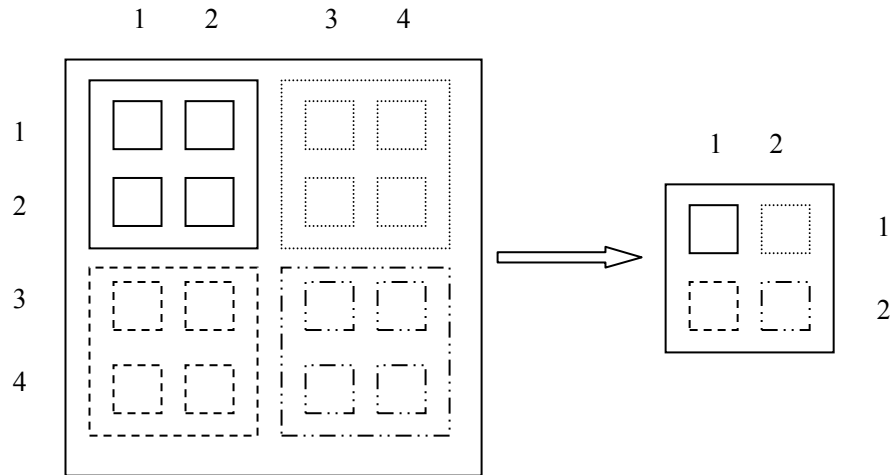


Fig. 4.5 An illustration of reducing the hologram size

Matrix reduction is based on the idea of reducing the size of the reconstructed hologram, as Fig. 4.5 shows. Suppose $U'_h(m', n')$, a shortened version of the reconstruction, is generated from the original reconstruction result $U_h(m, n)$. Each pixel of $U'_h(m', n')$ is the result of averaging square regions of side length p in $U_h(m, n)$. Thus, if the pixel numbers of the

image $U_h(m, n)$ is $N_x \times N_y$, the pixel numbers of the reduced image $U'_h(m', n')$ will be $\text{int}(N_x/p) \times \text{int}(N_y/p)$, with int standing for the integer part and p the reduction order. Each pixel of $U'_h(m', n')$ is obtained from the weighted average of the local neighbourhood of the pixel in the input $U_h(m, n)$, which can be represented mathematically as:

$$U'_h(m', n') = \frac{1}{p^2} \sum_{i=1}^p \sum_{j=1}^p U_h([p \cdot (m' - 1) + i], [p \cdot (n' - 1) + j]), \quad p = 1, 2, 3, \dots$$

$$m' = 1, 2, \dots, \text{int}(N_x/p), \quad n' = 1, 2, \dots, \text{int}(N_y/p) \quad (4.1)$$

This reduction can be understood to be a modified and localized low-pass filtering operation that generates a smaller output image in which all its entries are smoothed versions of corresponding square regions of size p . The bigger the reduction factor p , the smaller the speckle noise, while the bigger detriment of the image itself. Therefore, there must be a compromise in choosing a proper value for p .

The median filtering involves a convolution using a conventional image processing median kernel filter. Since speckle noise can be considered to be randomly occurring impulses, and a median filter is a good tool to remove binary or impulsive noise, whilst preserving spatial resolution, median filtering is quite effective in reducing the speckle noise in the reconstructed results. Like the choice of the reduction factor p that was addressed above, the width of the median filter window must also be optimised. Too small a filter width results in poor noise reduction, whereas too wide a filter results in blurring of the image.

It is more effective to reduce speckle noise by combining these two methods rather than using them independently. The order of implementing them in the processing is significant, resulting in different image qualities. This can be verified by considering the results presented in Fig. 4.6. Also it shows that the choices of reduction factor and the width of the median filter significantly affect the results as stated above. However, a fully proven formula that guarantees success in the speckle noise reduction process has hardly been established. As in many areas of digital image processing, it has instead been addressed by rule of thumb. Regardless of this lack of a strict formula, the digital manipulation of holograms in digital holography brings great flexibility and greater versatility to components to formerly purely optical methods, such as the application of digital image

processing methods here and the elimination of the zero-order term and the twin image by spatial filtering in Chapter 3.

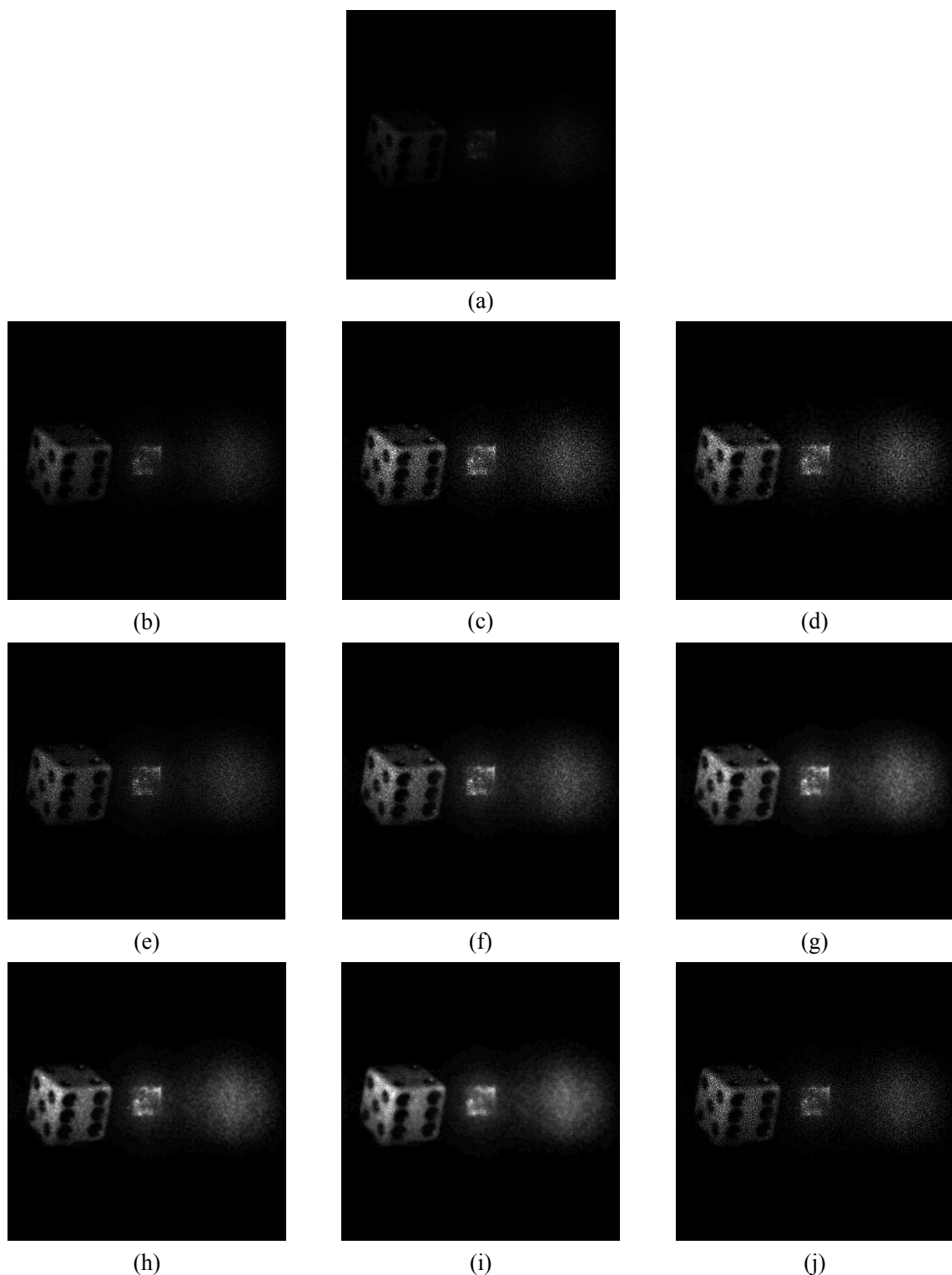


Fig. 4.6 The reconstruction results with reduced speckle noise by use of digital image processing: (a) an original reconstructed result of a die; matrix reduction of (b) 2×2 , (c) 4×4 , (d) 5×5 ; median filtering with a window width of (e) 3×3 , (f) 7×7 , (g) 9×9 ; (h) 2×2 matrix reduction followed by 5×5 median filtering; (i) 3×3 matrix reduction followed by 5×5 median filtering; (j) 5×5 median filtering followed by 3×3 matrix reduction.

4.3.2 Multiple holograms

The contrast value of a speckle pattern can be used to evaluate the effect of speckle noise. Speckle patterns with high contrast values contain more speckle noise than those with lower contrast values. The contrast of a speckle pattern ν is defined as

$$\nu = \sigma / \bar{I} \quad (4.2)$$

where \bar{I} is the mean value and σ is the standard deviation of the intensity distribution $I(m, n)$, both of which are described as

$$\bar{I} = \frac{\sum_{n=1}^N \sum_{m=1}^M I(m, n)}{N \times M}, \quad \sigma = \sqrt{\frac{\sum_{n=1}^N \sum_{m=1}^M [I(m, n) - \bar{I}]^2}{N \times M}} \quad (4.3)$$

Here $N \times M$ is the size of the speckle pattern. The contrast value of a standard speckle pattern is always equal to 1 because of its intensity distribution with a negative exponent.

Suppose k speckle patterns are represented as I_i , $i = 1, 2, \dots, k$. They follow the same intensity distribution but are statistically independent of each other. The mean value of I_i is denoted by \bar{I}_i , and the standard deviation of I_i is denoted by σ_i . Therefore the contrast value of I_i is

$$\nu_i = \sigma_i / \bar{I}_i \quad (4.4)$$

The averaged pattern (\hat{I}) of these speckle patterns is calculated by

$$\hat{I} = \sum_{i=1}^k I_i / k \quad (4.5)$$

So the mean value of \hat{I} should be equal to \bar{I}_i :

$$\bar{\hat{I}} = \bar{I}_i \quad (4.6)$$

Meanwhile, the relation between the standard deviation ($\hat{\sigma}$) of \hat{I} and σ_i is

$$\hat{\sigma} = \sigma_i / \sqrt{k} \quad (4.7)$$

According to Eqs. (4.2), (4.3), (4.6) and (4.7), the contrast value ($\hat{\nu}$) of \hat{I} is calculated by

$$\hat{v} = \hat{\sigma} / \hat{I} = v_i / \sqrt{k} \quad (4.8)$$

It is obvious in Eq. (4.8) that the contrast value of an averaged speckle pattern, taken from a series of k speckle patterns, is reduced by a factor of $1/\sqrt{k}$ compared with that of a single speckle pattern. Hence speckle noise can be significantly reduced by averaging a number of speckle patterns that do not correlate with each other. Because a hologram is a recorded interference pattern of the combined object and reference waves, it is therefore eligible to apply this principle to holographic applications.

The experimental setup is usually based on a typical off-axis holographic geometry, as shown in Fig. 4.7. However different researchers have used different methods to generate the series of holograms, which must have the same intensity distribution but uncorrelated speckle patterns.

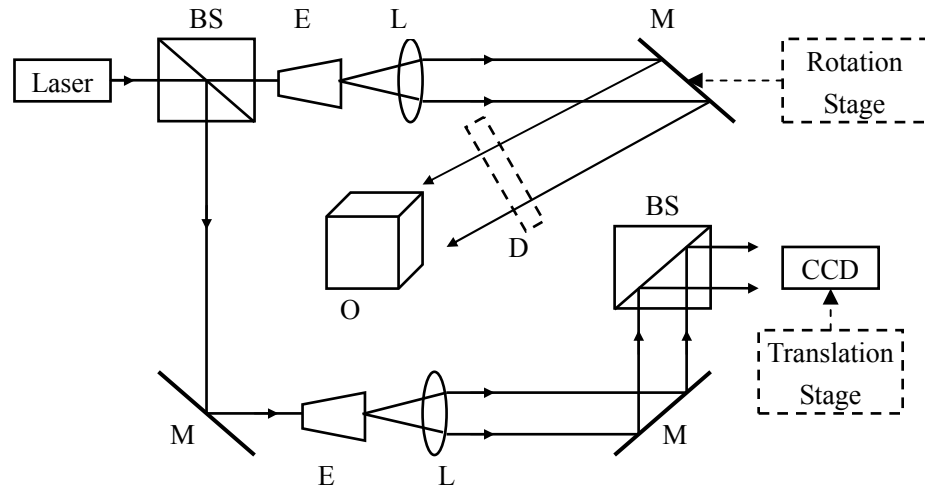


Fig. 4.7 The experimental setup to record multiple holograms to reduce speckle noise (BS, beamsplitter; E, objective; L, collimating lens; M, mirror; O, object; D, diffuser).

It is quite simple to couple a diffuser into the geometry in order to introduce the change in the speckle pattern. Most of the approaches (Dainty, 1975, Lowenthal and Joyeux, 1971, Dubois et al., 2004) to reduce the undesired effects of speckle noise in the coherent imaging system are based on moving diffusers during the integration time of the detector. The speed of the diffuser motion is tuned so that the phase of the illumination is temporally varying faster than the temporal resolution of the detector, providing an illuminating beam with reduced spatial coherence and consequently reduced speckle noise. However Garcia-Sucerquia et al. (2006) proposed a method which does not have to move the

diffuser within the integration time of the detector. Moreover, the speed and the movement of the diffuser are not necessarily as strictly controlled as in the former methods which makes the system simpler and easier to implement. It is an off-axis holographic system, but with a diffuser placed before the object to affect the speckle pattern of the illumination wave. The diffuser is shown as a dashed line in Fig. 4.7. The diffuser is kept still during the integration time of the detector, so that spatial coherence is preserved. This can be verified in Fig. 4.8(b) showing that the hologram recorded with a diffuser is reconstructed with high-contrast speckle noise, which is similar to the hologram recorded without a diffuser shown in Fig. 4.8(a). Fig. 4.8(b) is darker than Fig. 4.8(a) because of the reduction in the intensity of the object wave that was caused by the scattering of the diffuser. Then another hologram is recorded by shifting the diffuser to another position, but ensuring that the other parts of the geometry remain in the same condition as the last exposure. This procedure is repeated several times in order to produce a series of holograms with different speckle patterns. It should be noted that the shifts are made by keeping the diffuser in the same plane and ensuring that different areas of it are illuminated in each position. Therefore, statistically similar but uncorrelated speckle patterns will illuminate the object. Such conditions allow the shifts to be performed longitudinally or by rotating the diffuser in steps greater than the average size of the scattering surface asperities. Thereafter, several reconstructed holograms are superimposed on an intensity basis to generate a result with a reduction in the level of speckle noise. In Fig. 4.8, a series of reconstructed amplitude-contrast pictures of a die are displayed. The die is the same one originally shown in Fig. 2.6(a). With greater increments of the number of the superimposed images, the speckle noise reduction was improved and greater object detail was clearly revealed. Comparing Fig. 4.8(a) and (b), the introduction of a diffuser makes the reconstructed image of the die more uniform. This is because of the uniform illumination of the die through the diffuser.

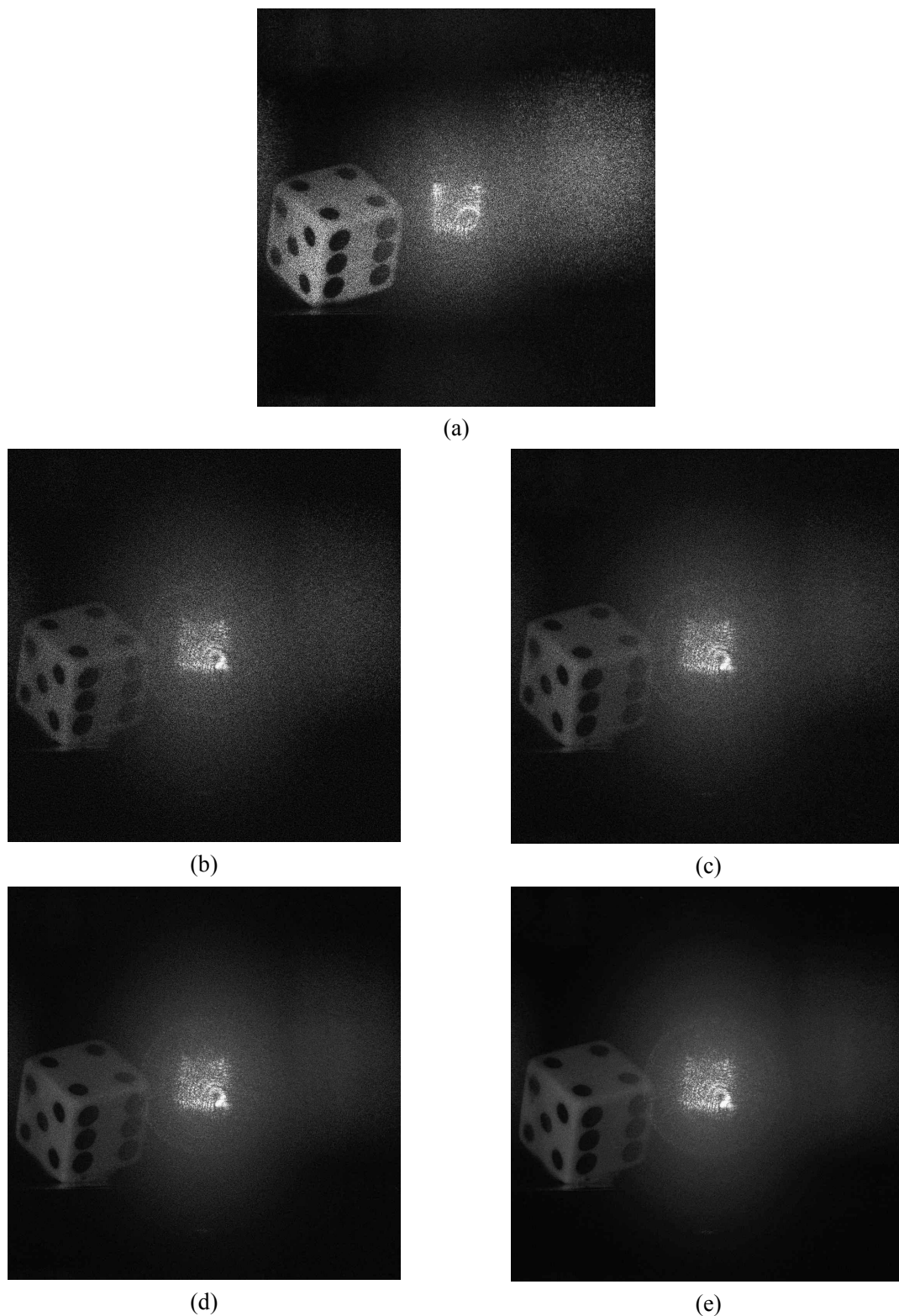


Fig. 4.8 The reconstructed amplitude-contrast images of a die : (a) without a diffuser; (b) with a diffuser in a particular still position; by superimposing (c) 2, (d) 6, (e) 12 holograms according to the method described above.

Baumbach et al. (2006) proposed a new method of obtaining the holograms of different speckle patterns by shifting the detector. The CCD camera was placed on an x - y translation stage (indicated in Fig. 4.7 by dashed lines) to enable a shift that was parallel to the surface

of the object. By changing the camera position laterally which assures the shift of the camera was parallel to the surface of the object, the speckle field of the reconstructed object was changed and this was then recorded by the camera. The linear phase factor caused by the camera shift can be calculated and deducted from the directly reconstructed results to ensure that each phase map was similar to the phase map in the reference position. Finally all the processed phase maps were averaged to achieve a result with less speckle noise and better spatial resolution by applying the following relation:

$$\varphi_{mean}(x, y) = \arctan \frac{\sum_{i=1}^N \sin[\varphi_i(x, y)]}{\sum_{i=1}^N \cos[\varphi_i(x, y)]} \quad (4.9)$$

where N is the number of images, φ_i is the phase value at the i th position.

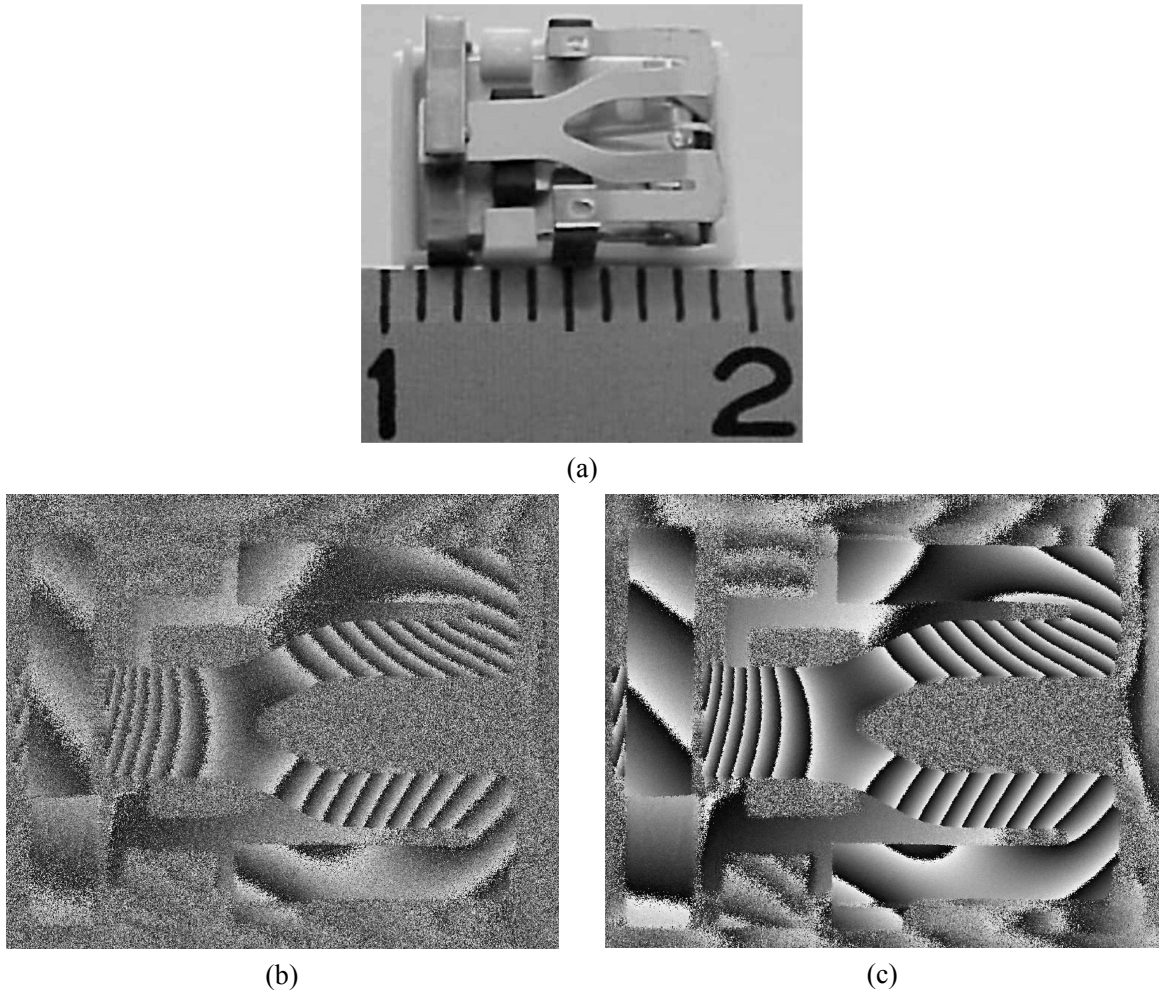


Fig. 4.9 (a) image of a relay; (b) result of a single deformation measurement of the relay; (c) result after averaging with 25 pictures (Reprinted with permission from Baumbach et al. (2006). Copyright(2006), Optical Society of America).

This method can actually be interpreted as a generation of a large synthetic aperture consisting of many small apertures given by the single CCD. The example presented in Fig. 4.9 shows that this technique can suitably reduce speckle noise and recover spatial resolution in digital holography. The removal of the linear phase factor caused by the translation of the CCD sensor ensures accurate phase measurement. However, the use of the x - y translation stage makes the experimental system quite complicated and expensive.

Quan et al. (2007) acquired multiple holograms by continuously changing the incidence angle of the wave used to illuminate the object. The mirror used to direct this illumination wave was mounted on a rotation stage that is shown in dashed lines in Fig. 4.7. By the rotation of this mirror, a set of holograms with different speckle patterns were obtained. These holograms were individually numerically reconstructed. By averaging the intensities of these reconstructed images, the speckle noise was drastically suppressed and the reconstructed result exhibited with a better contrast and clarity. This can be viewed from the results from the die shown in Fig. 4.10.

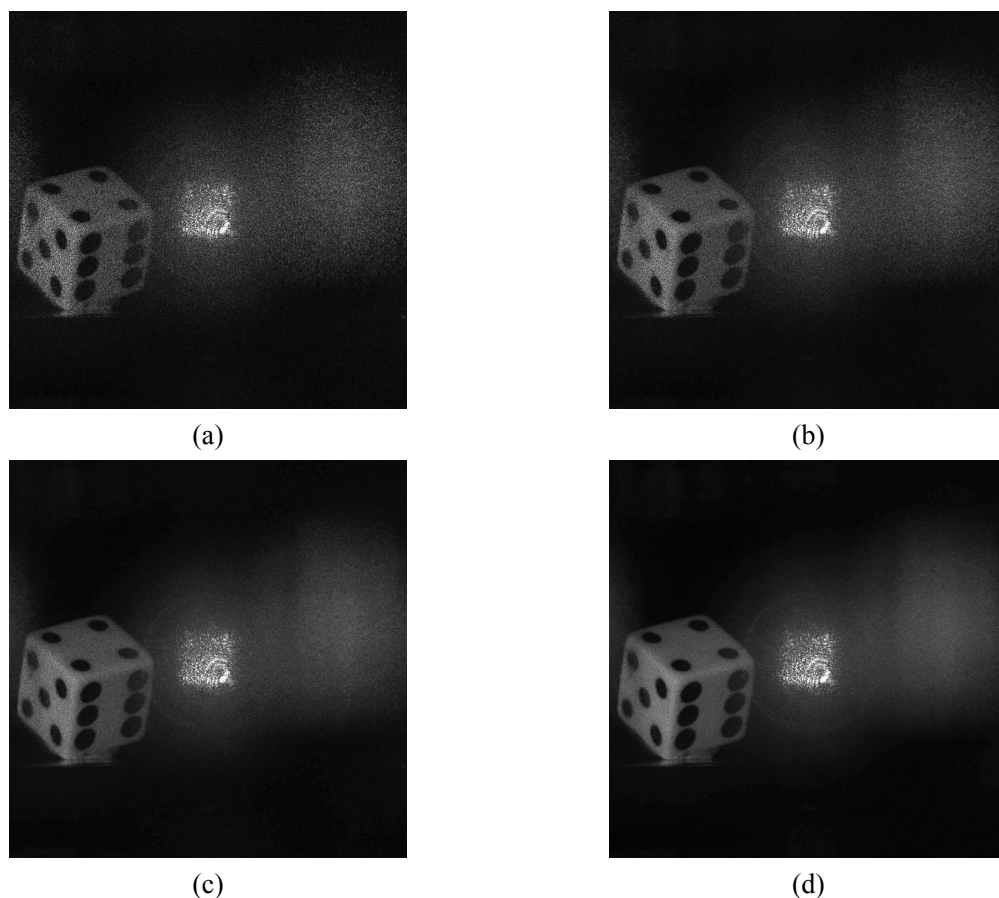


Fig. 4.10 Reconstructed amplitude-contrast images of the die obtained by: (a) 1, (b) 2, (c) 6, (d) 12 holograms with different angles to illuminate the die controlled by a rotation stage

The results improve with the use of increasing numbers of holograms, but this is limited by the size and intensity of the optical field on the object. Therefore this method is simple and suitable for smaller objects. It can also be implemented so as to perform shape and deformation measurements.

For the methods using multiple holograms, the improvement factor of contrast value does not strictly follow a relation of $1/\sqrt{k}$. That is because the images reconstructed for every single hologram partially correlate with each other and the correlation coefficients are distributed randomly around 0, but are not exactly equal to 0. The situation varies under different experimental conditions and only can be evaluated experimentally.

4.3.3 Deconvolution of the aperture function

This method was recently proposed by Cai and Wang (2008). Through a careful investigation of the whole process of digital holography, including recording and reconstruction, they came to the conclusion that the speckle noise is aggravated after diffraction by a window function; moreover, the speckle noise is more aggravated by smaller window widths. Therefore the speckle noise in the reconstructed image of hologram can be reduced by reducing the aggravated noise due to the small-aperture diffraction through deconvolution.

Suppose the object surface consists of N points and the complex amplitude reflectivity of each point is $U_o(\xi, \eta)$. The coordinate system is shown in Fig. 2.3. Then the reconstructed real image on the image plane $U_{ir}(\xi', \eta')$ can be described as

$$U_{ir}(\xi', \eta') = |U_r(x, y)|^2 \exp\left[i \frac{2\pi}{\lambda d_i} (\xi'^2 + \eta'^2)\right] \sum_n U_o(\xi, \eta) \exp\left[-i \frac{2\pi}{\lambda d} (\xi^2 + \eta^2)\right] \times \int_{\Sigma} \int \exp\left[-i 2\pi \left(\frac{\xi' - \xi}{\lambda d} x + \frac{\eta' - \eta}{\lambda d_i} y\right)\right] dx dy \quad (4.10)$$

where $U_r(x, y)$ represents the reference wave incident onto the hologram plane, λ is the wavelength of the reconstruction light, Σ is the area of the recording aperture on the hologram plane, d is the distance between the object and hologram plane, and d_i is the distance between the hologram plane and the image plane. If $d = d_i$, i.e. the reconstructed

image is at the same distance from the hologram plane as the original object, Eq. (4.10) can be written as

$$U_{ir}(\xi', \eta') = |U_r(x, y)|^2 \exp\left[i \frac{2\pi}{\lambda d} (\xi'^2 + \eta'^2)\right] \sum_n U_o(\xi, \eta) \exp\left[-i \frac{2\pi}{\lambda d} (\xi^2 + \eta^2)\right] \times P\left(\frac{\xi' - \xi}{\lambda d}, \frac{\eta' - \eta}{\lambda d}\right) \quad (4.11)$$

where $P\left(\frac{\xi' - \xi}{\lambda d}, \frac{\eta' - \eta}{\lambda d}\right)$ is the Fourier transformation of the hologram aperture.

Therefore it is evident that the reconstructed image of the hologram is the convolution of the original object's reflected light distribution when illuminated by coherent light and the Fourier transformation of the hologram aperture function. Assuming that the hologram aperture is a rectangle, with dimensions a and b in width and length, the aperture function can be expressed as

$$p(x, y) = \text{rect}(x/a) \text{rect}(y/b) \quad (4.12)$$

and its Fourier transformation is

$$P(X, Y) = \frac{ab}{\lambda d} \text{sinc}\left(\frac{aX}{\lambda d}\right) \text{sinc}\left(\frac{bY}{\lambda d}\right) \quad (4.13)$$

Therefore, a deconvolution process can be applied to the Fourier transformation of the aperture function in order to reduce speckle noise in the reconstructed image that is caused by diffraction through the hologram aperture. An example is shown in Fig. 4.11.

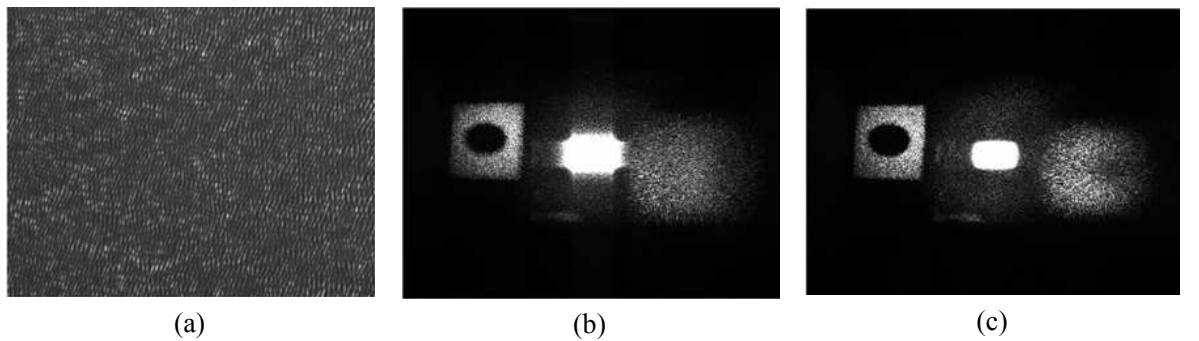


Fig. 4.11 (a) The digital hologram of a circle; (b) its reconstruction without deconvolution and (c) after deconvolution (Reprinted from (Cai and Wang, 2008). Copyright (2008), with permission from Elsevier)

From the example it can be seen that the speckle noise from the aperture in the zero-order diffraction term has been reduced. Although this kind of noise is diffracted into every order

of the diffraction term, it has less influence on higher diffraction orders than on the zero-order diffraction term. That is why the image of the object is not improved as much as the zero-order term in Fig. 4.11(c).

4.4 Resolution improvement

As stated in section 1.5.1, the size of the object is limited by Δx , the distance between neighbouring pixels of CCD sensor. It is described by Eq. (1.9):

$$\theta_{\max} = 2 \arcsin\left(\frac{\lambda}{4\Delta x}\right) \approx \frac{\lambda}{2\Delta x} \quad (4.14)$$

where θ_{\max} is the maximum resolvable angle between the object wave and the reference wave, which restricts the size of the object, and λ is the recording wavelength. This equation is valid for small angles. This relation has to be carefully maintained in the recording procedure. Otherwise, the contrast of the whole hologram decreases or even vanishes in extreme cases. Fig. 4.12 shows the effect of different Δx values for the CCD sensor in digital holography. The object is a die with side length of about 13mm that was originally shown in Fig. 2.6(a). Only the effective resolution of the CCD was changed by subsampling in these images in order to vary Δx , but all other conditions remained unchanged. The die in Fig. 4.12(a) is shown clearly, but the image shown in Fig. 4.12(b) is blurred and incomplete, caused by aliasing due to insufficient samples. With a further reduction in the resolution of the CCD, the fringe pattern in the hologram cannot be resolved by the CCD so that the image of the die disappears in Fig. 4.12(c).

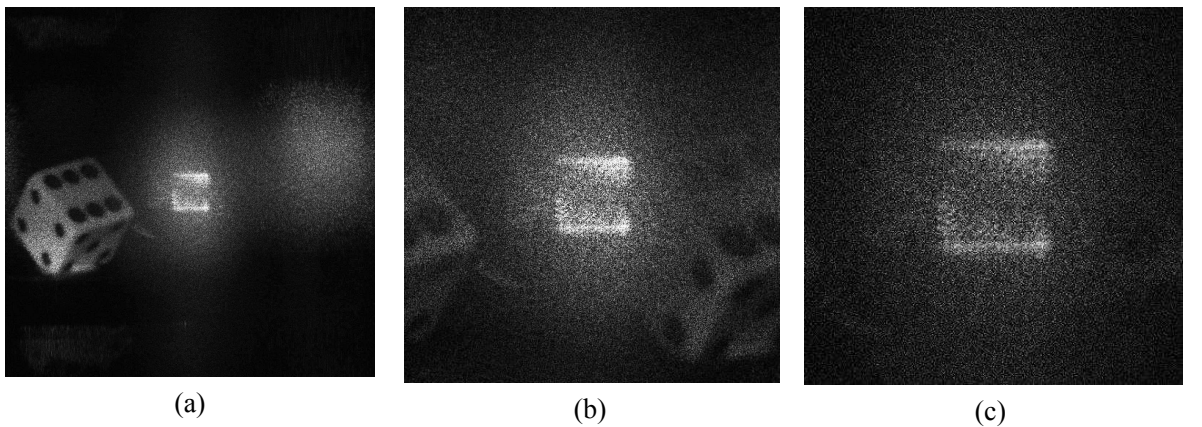


Fig. 4.12 The reconstruction results of a die with different resolutions of the CCD sensor: (a) 1024×1024, equivalent to a neighbouring pixel distance of 6.45 μm ; (b) 512×512, equivalent to a neighbouring pixel distance of 12.9 μm ; (c) 340×340, equivalent to a neighbouring pixel distance of 19.35 μm .

According to Eq. (4.14), either a laser source with a long wavelength or a CCD camera with a very small distance between neighbouring pixels is able to enlarge the maximum resolvable angle. In practice, most laser sources with long wavelengths are diode lasers, which have a large divergence angle and poor coherence. Meanwhile wavelengths above 750 nm are invisible to human eyes and normal CCD cameras also have insufficient quantum efficiency in this wave band. Therefore, most researchers use light sources such as HeNe lasers, Argon Ion lasers and high powered solid-state pulse lasers. It is straightforward to achieve better spatial frequencies by reducing the pixel size of the CCD sensor. However, the light intensity falling on a single pixel drops when the pixel size decreases. This generates shot noise that severely degrades the image quality. There is obviously a restriction in the amount by which pixel size may be reduced, whilst remaining free of the degrading effect caused by shot noise. Current image sensor technology has almost reached this level of $40\text{ }\mu\text{m}^2$ (Park et al., 2003). Since both of these direct methods for increasing object size are impractical to apply, other methods have to be found to measure objects of larger sizes at a reasonable recording distance.

Schnars et al. (1996) proposed a simple method to record large objects using digital holography. The experimental geometry is shown in Fig. 4.13. A divergent lens is arranged between the object and the target. This lens generates a reduced virtual image of the object at a distance d' . The wave emerging from this virtual image is the object wave and interferes with the reference wave to generate the hologram. This method is able to measure large objects at reasonably short distances. But the lens has the property of spatial filtering and causes an undesired optical bandlimiting effects, so that some object detail is lost.

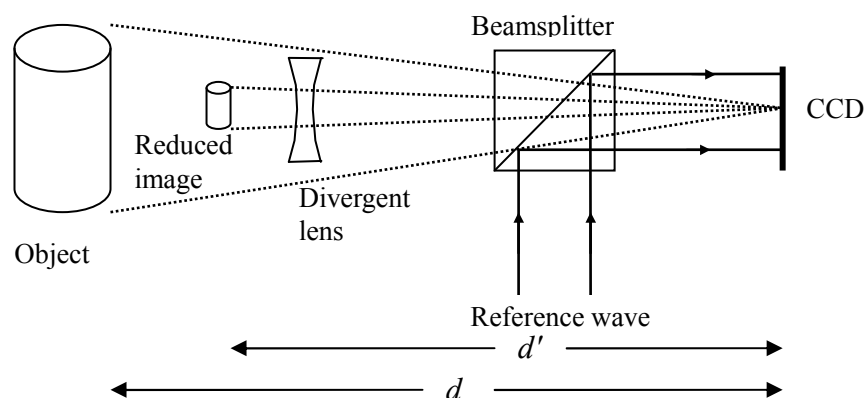


Fig. 4.13 Recording geometry for large objects using a divergent lens

Because of the limited size of commercial CCD sensors, which is typically about $6 \times 8 \text{ mm}^2$, only a very small part of the object wave can be received by the sensor. It is hard to improve the resolution by enlarging the sensitive area of the CCD because it leads to an increase in capacitance, which results in a low charge transfer rate. However it can be improved by making more of the object wave reach the CCD chip. It is well known that light incident upon a grating often splits into three beams. One of these beams will propagate along the original direction, and the other two beams will deviate from the original direction with an angle of $\pm \lambda/\Delta$, where λ and Δ are the illumination wavelength and the period of the grating, respectively. This situation is depicted in Fig. 4.14:

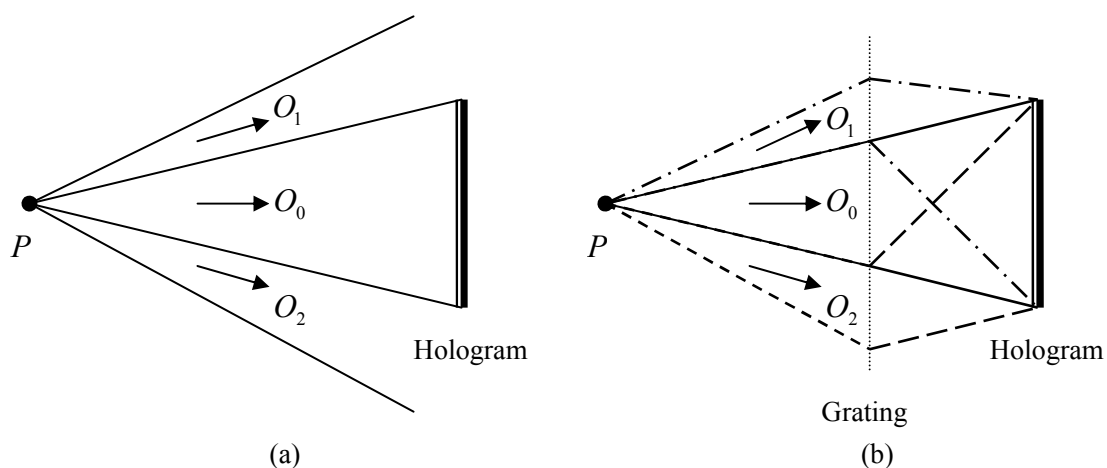


Fig. 4.14 Ray diagrams of the object wave: (a) without grating; (b) with a grating.

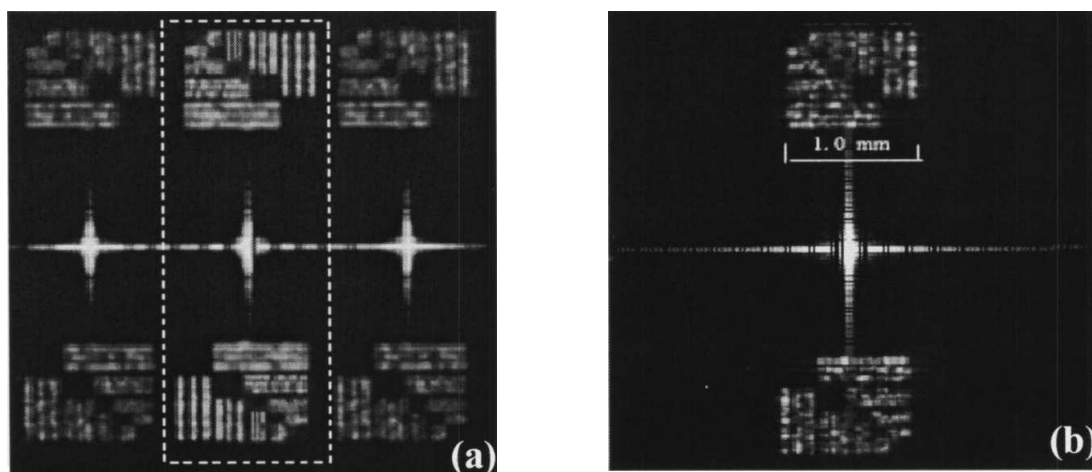


Fig. 4.15 Reconstruction results (a) using a grating to achieve super-resolution and (b) normal method without using a grating (Reprinted with permission from (Liu et al., 2002). Copyright (2002), American Institute of Physics).

It is obvious that more of the object wave will reach the sensitive area of the CCD with the use of a grating. Liu et al. (2002) noticed this property and inserted a grating into the off-axis hologram recording geometry. A grating was positioned between the object and the CCD sensor, where it does not affect the reference wave. In the numerical reconstruction, the function of the grating has to be considered in order to get the exact object wave in the object plane. This adds some complexity to the numerical reconstruction. Their measurement of a resolution target shows a noticeable improvement in spatial resolution as shown in Fig. 4.15.

Although the practical problem of increasing the sensitive area of CCD is hard to overcome with contemporary technology, an approach based on this idea, which is called a synthetic aperture, has been successfully employed in radio and optical astronomy. Two or more separated sensor arrays are used to record the same scene and a combination of the recorded holograms is used to generate a new hologram with greater pixel numbers. It is quite easy to understand how this method achieves the improved spatial resolution when applied to holography. Every part of a hologram contains information from the entire object. Hence the resolution of the reconstructed image improves with an increase in the effective pixel numbers. However in the optical hologram reconstruction, the increase in the number of pixels produces a reduction of the speckle size due to the enlargement of the aperture. The concept of a synthetic aperture can be realized by various methods in digital holography. Binet et al. (2002) presented an active synthetic aperture-imaging system which is a combination of a synthetic aperture and phase-shifting digital holography. Multiple digital holograms are generated by rotating the object and are overlapped according to the estimation and compensation of their relative positions through a speckle cross-correlation algorithm. Massig (2002) overlapped multiple holograms obtained by moving the camera linearly. Kreis and Schluter (2007) captured two holograms by two cameras simultaneously, then embedded both holograms in an artificial large hologram using zero padding of pixels with no grayscale value. In this method, the shift between two single holograms has to be determined to subpixel accuracy. An example of the kind of super-resolution result obtained by use of synthetic aperture is given in Fig. 4.16.

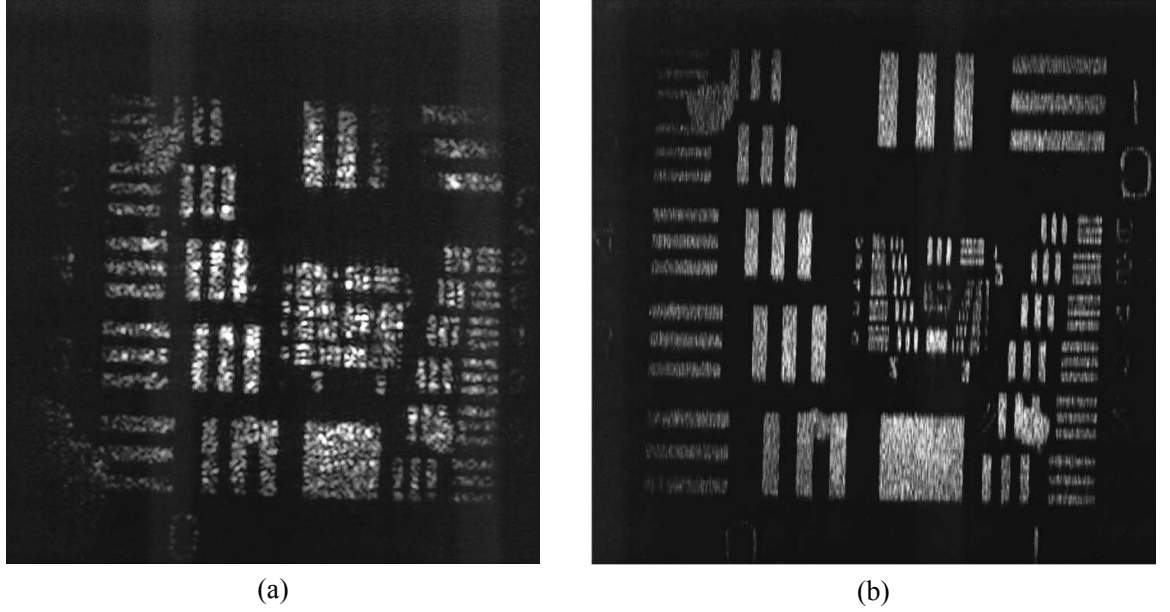


Fig. 4.16 Synthetic aperture image of an USAF scattering test target computed with (a) a single hologram 128×128 pixels; (b) a synthetic aperture (2048×128) composed of 33 subpupil holograms merged coherently (Reprinted with permission from Binet et al. (2002). Copyright(2002), Optical Society of America).

In fact, the hologram is the aperture of the recording system. According to the theory of diffraction, the size of the reconstructed pixel in the Fresnel approximation given in Eq. (2.25) is

$$\Delta\xi = \frac{\lambda d}{N\Delta x} \quad (4.15)$$

which is the half diameter of the airy disk or speckle diameter in the plane of the reconstructed image, which limits the optical resolution. However, increasing the aperture size by using more than one CCD does not automatically improve the image resolution because the larger synthetic aperture requires a longer recording distance which is illustrated in Fig. 4.17. θ_{\max} is restricted by Δx according to Eq. (4.14). The pixel size of most commercial CCDs is of the order of $5 \mu m$, hence the object has to be set at a longer distance in order to perform synthetic aperture methods.

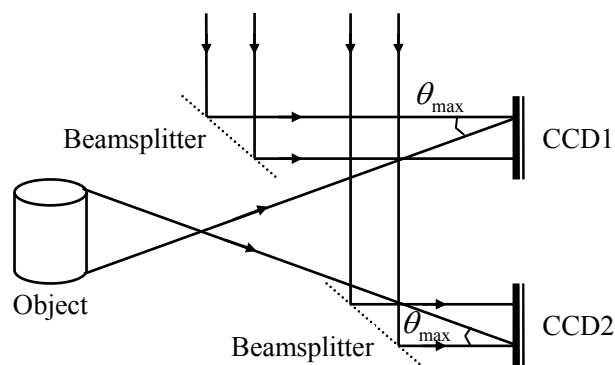


Fig. 4.17 Scheme showing the use of two CCDs to create a synthetic aperture (redrawn from (Schnars and Juptner, 2005))

As discussed previously, the physical pixel size of CCD sensors seems unlikely to be significantly further reduced using present technology. Meanwhile, the idea of increasing the size of the sensor chip is also invalid because this results in an increase in chip capacitance which leads to a slow charge transfer rate. In order to overcome these restrictions in obtaining higher resolution images than those provided by CCD, a signal-processing based approach known as super resolution (SR) image reconstruction provides an alternative. The term SR was originally used in optics, and it refers to algorithms that operate mainly on a single image to extrapolate the spectrum of an object beyond the diffraction limit, i.e. SR restoration (Kang and Chaudhuri, 2003). The synthetic aperture method mentioned above is a type of SR restoration. These two SR concepts, SR image reconstruction and SR restoration, have a common focus in recovering high-frequency information that has been lost or degraded during image acquisition. However, the cause of this loss of high-frequency information differs between these two concepts. SR restoration in optics attempts to recover information beyond the diffraction cutoff frequency, while the SR image reconstruction method used in engineering tries to recover high-frequency components that have been corrupted by aliasing. In the rest of this chapter, the term SR is used to specify SR image reconstruction.

The SR image is produced from multiple low-resolution (LR) images. The basis for increasing the spatial resolution in SR techniques is to capture multiple LR images from the same scene. This can be achieved by several acquisitions from one camera, or from multiple cameras installed in different positions. Each LR image represents a different “view” of the same scene. These LR images are subsampled (aliased) and shifted with subpixel precision. They cannot be shifted by integer units of pixels because that would

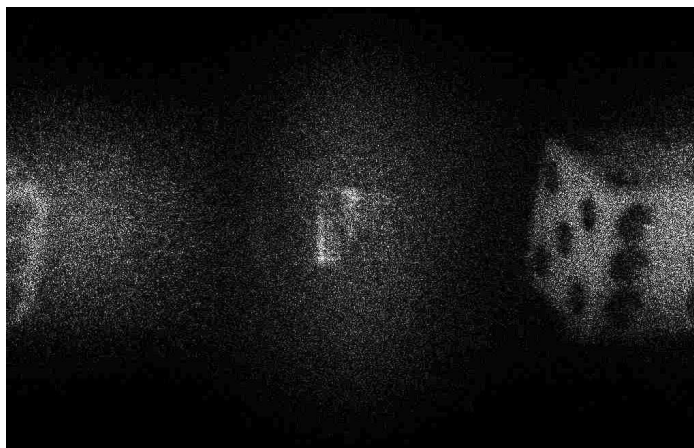
cause each image to contain the same information, thus there would be no new information that could be used to reconstruct an SR image. However, using different subpixel shifts for multiple captures ensure that new information is obtained in the LR images and each LR image cannot be obtained from the others. In this case, the new information contained in each LR image can be exploited to obtain an SR image.

Most of the SR image reconstruction methods consist of three stages: registration, interpolation, and restoration (Park et al., 2003). These steps can be implemented separately or simultaneously, according to the reconstruction methods adopted. The estimation of motion information is referred to as the registration stage, and here the relative shifts between the LR images when compared to the reference LR image are estimated with fractional pixel accuracy. Obviously, accurate subpixel motion estimation is a very important factor in the success of the SR image reconstruction algorithm. In practice, the motions between the LR images are generated by the controller in the imaging system. Therefore high accuracy of the controller is essential to the performance of SR image reconstruction. Image interpolation has been extensively investigated in order to increase the number of effective pixels in a single image. The quality of an image magnified from an aliased LR image is inherently limited, even though ideal *sinc* basis functions are employed. This is because single image interpolation cannot recover the high-frequency components lost or degraded during the LR sampling process. For this reason, image interpolation methods are not considered as SR techniques. To achieve further improvements in this field, the next step requires the utilization of multiple data sets in which additional data constraints can be used from several observations of the same scene. The fusion of information from various observations of the same scene allows the SR reconstruction of the scene. Image restoration based on image processing rather than on the type of optical restoration mentioned previously is also related to SR image reconstruction. The goal of image restoration is to recover a degraded (e.g., blurred, noisy) image, but it does not change the size of image.

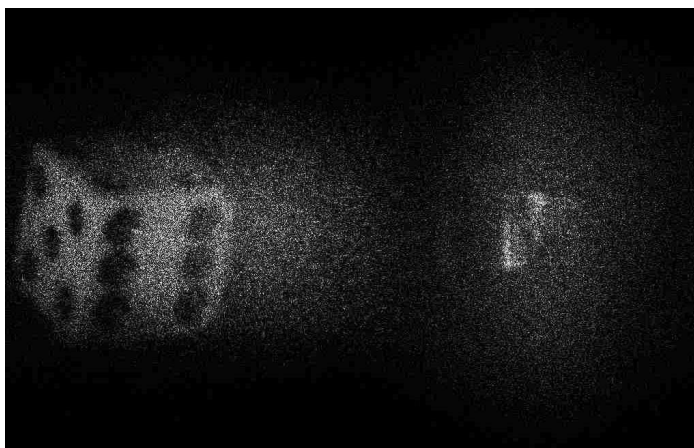
In our experiments, a ProgRes[®] MF^{scan}, CCD camera with a “microscanning” function, which is also sometimes referred to as “pixel shifting”, was used to capture SR digital holograms via subpixel movements. The image sensor of such a camera can be moved with sub-pixel displacements by means of piezoelectric actuators. Using this technique the image plane is “scanned” by the CCD array sensor in a similar manner to a line scanner.

The successively captured single images are then recombined to produce an image with greatly increased spatial resolution.

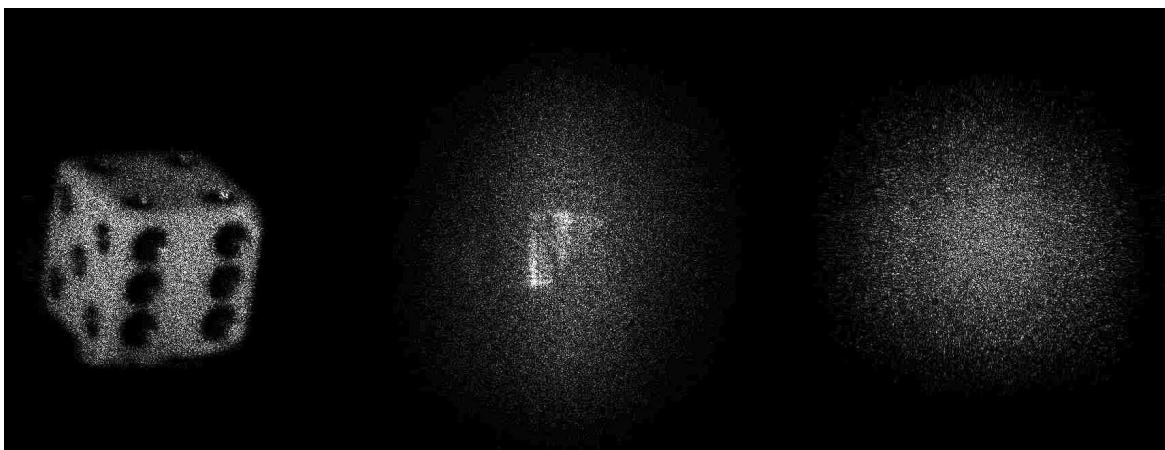
Fig. 4.18 shows experimental results using the ProgRes[®] MF^{scan} camera, with a die as the object. The hologram recording geometry is an off-axis set-up as shown in Fig. 4.1(b). A 35mw HeNe laser with a wavelength of 632.8nm was used as the light source. The size of the die was about 13mm×13mm×13mm, and the distance between the front of the die and CCD sensor chip was 604mm. The CCD chip has a resolution of 1360×1024 but only the central 1024×1024 pixels are used to perform numerical reconstruction. The size of each pixel in the CCD is 6.45μm×6.45μm. According to Eq. (4.15), the maximum size of object that would be resolvable is 29mm. But in off-axis holography, the introduction of an additional carrier frequency by the reference wave and the existence of the twin image reduce the resolvable resolution of the system to less than half of the maximum resolution. This is revealed in Fig. 4.18(a) which suffers from severe aliasing. The display of the die is incomplete and misplaced, because the image of the object should be on the left of the zero-order image, according to its original recording setup. If an angle is introduced to the simulated reference wave in the numerical reconstruction that is different from that used in the recording process, the carrier frequency is changed and the position of the image of the object and zero-order image move accordingly. This can be seen clearly in Fig. 4.18(b). Although the image of the object is now complete, the aliasing is still obvious and the twin image is superposed with the image of the object. The quality of Fig. 4.18(c) is highly improved compared to the former Fig. 4.18(a) and (b). The aliasing vanishes and the contrast of the image is much better. It is reconstructed from a hologram taken in 4-scan mode. By shifting the CCD chip three times, by a distance of half the pixel size in *x*- and *y*-dimensions, four holograms with resolutions of 1024×1024 are recorded and combined to produce a single super resolution hologram with a resolution of 2048×2048. This verifies the ability of the SR image reconstruction technique to increase the resolution of holograms and eliminate the aliasing effects caused by undersampling. Therefore, larger objects can be recorded and the distance between the object and the hologram can be shortened without resorting to the use of any optical lens.



(a)



(b)



(c)

Fig. 4.18 The numerical reconstruction results for the case when the recording conditions are such that the angle between the die and the reference wave is bigger than that which the CCD camera can resolve: (a) normal reconstruction; (b) normal reconstruction with an angle introduced for the reference wave, which enables the display of the die completely; (c) reconstruction of a super hologram with a resolution of 2048×2048 captured by 4-scan mode.

Using the same die, placed at an angle that can be resolved by the camera, the reconstruction results are shown in Fig. 4.19. The zero-order term and the twin image are

cropped, and only the desired image of the die is displayed. Comparing these pictures, the one produced using 9-scan mode shows greater detail and has better contrast.

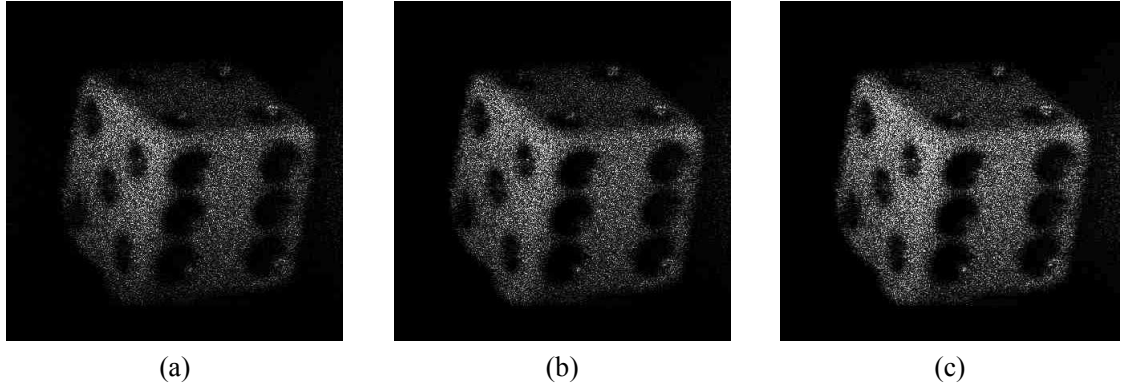


Fig. 4.19 The intensity-contrast reconstructed die with different recording modes: (a) 1-scan; (b) 4-scan; (c) 9-scan.

There are different methods to combine the LR images as a SR image (Kornis and Gombkoto, 2005, Baldi, 2006). They vary from a simple, weighted summation to more complex methods in the frequency space. The family of interleave methods are the simplest and the most cost-effective in the calculation. They can be described by the general formulation:

$$I_k = \sum_{l=1}^N w_{k,l} J_l \quad (4.16)$$

where I_k is the k^{th} SR pixel, J_l is the l^{th} LR pixel and $w_{k,l}$ weights the contribution of the LR pixel l to the SR pixel k . The methods differ in the manner in which $w_{k,l}$ is chosen. For the interlace method, every input LR pixel falls only on one output SR pixel so that every output SR pixel gets information from one input LR pixel only, and not from other input LR pixels. For the four-quadrant 2×2 drizzle method, every input LR pixel is equally shared between four output SR pixels.

The schematic diagrams of these two methods are shown in Fig. 4.20. Four LR images I_k ($k = 1, 2, 3, 4$), each with pixel number of $m \times n$, are used to generate a SR image. They are captured by moving the camera by half of the distance between two neighboring pixels in both x and y axes which are displayed in Fig. 4.20 with number 1, 2, 3 and 4. Suppose the output SR image J is of pixel number of $2n \times 2m$. Then for the interlace method, the output SR pixels are given

$$J(2x, 2y) = I_1(x, y) \quad (4.17)$$

$$J(2x+1, 2y) = I_2(x, y) \quad (4.18)$$

$$J(2x, 2y+1) = I_3(x, y) \quad (4.19)$$

$$J(2x+1, 2y+1) = I_4(x, y) \quad (4.20)$$

While for the four-quadrant 2×2 drizzle method, the output SR pixels are given

$$J(2x, 2y) = I_1(x, y) + I_2(x-1, y) + I_3(x, y-1) + I_4(x-1, y-1) \quad (4.21)$$

$$J(2x+1, 2y) = I_1(x, y) + I_2(x, y) + I_3(x, y-1) + I_4(x, y-1) \quad (4.22)$$

$$J(2x, 2y+1) = I_1(x, y) + I_2(x-1, y) + I_3(x, y) + I_4(x-1, y) \quad (4.23)$$

$$J(2x+1, 2y+1) = I_1(x, y) + I_2(x, y) + I_3(x, y) + I_4(x, y) \quad (4.24)$$

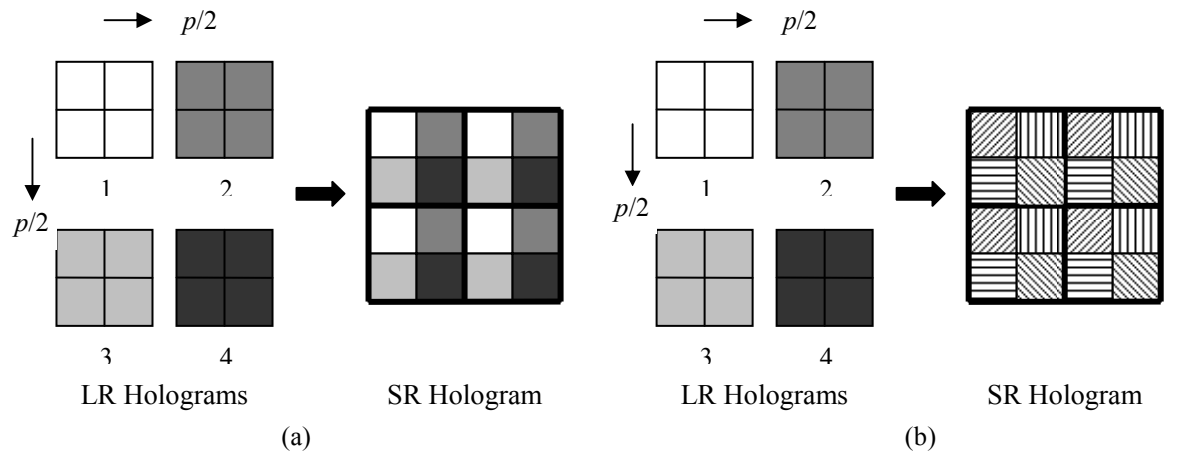
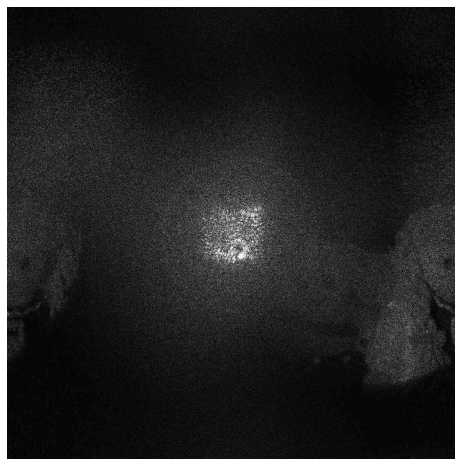
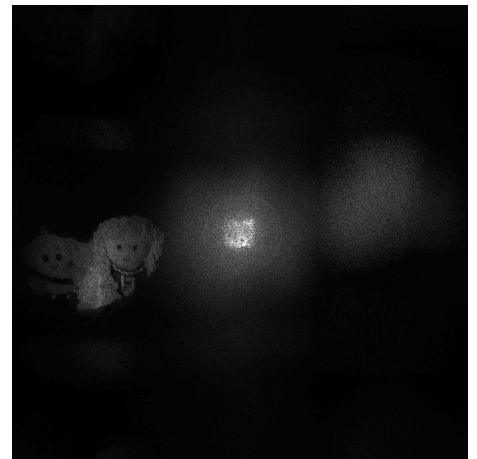


Fig. 4.20 Schematic diagrams of interleave methods: (a) the interlace method; (b) four-quadrant 2×2 drizzle method. (p is the distance between two neighboring pixels.)



(a)



(b)

Fig. 4.21 The reconstruction results of a small statue obtained from: (a) a LR hologram; (b) a SR hologram generated by ProgRes[®] MF^{scan}.

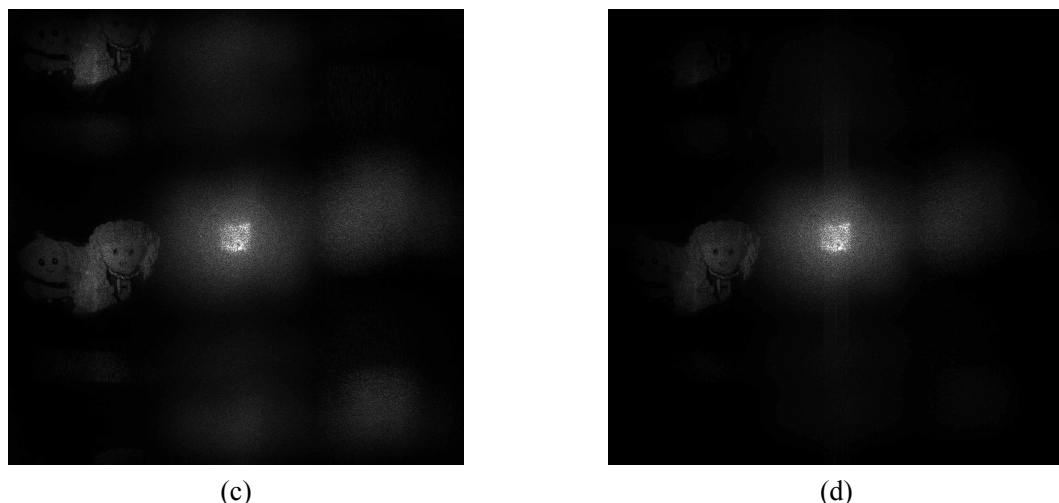


Fig. 4.21 (Continued) (c) a SR hologram generated by the interlace method; (d) a SR hologram generated by the four-quadrant 2×2 drizzle method.

From the results shown in Fig. 4.21, it is clear that the reconstruction results are affected by different methods used to generate a SR hologram. The actual algorithm used in the ProgRes[®] MF^{scan} camera is commercially confidential and the company has refused to disclose the details. But the result obtained by this method in Fig. 4.21(b) shows that this method provides the best reconstruction results in our experiments. Therefore, the SR holograms directly obtained from the ProgRes[®] MF^{scan} are used in the optical contouring process of a 3D object described in Chapter 5.

It is natural to use interpolation to increase the size of the hologram. However, as was explained previously, single image interpolation is not able to recover high-frequency components that were lost or damaged during the LR sampling process. This is shown in Fig. 4.22. A hologram of the same die is recorded at the distance of 629mm with the same optical system to obtain Fig. 4.18. It is shown in Fig. 4.22(a) with a resolution of 1024×1024. The amplitude-contrast image reconstructed from it is displayed in Fig. 4.22(b), where the three terms are separated well. Then the original hologram is downsampled with a new resolution of 512×512. Every other pixel in x - and y - dimensions is preserved to form the downsampled new hologram. The amplitude-contrast image reconstructed from this downsampled hologram is shown in Fig. 4.22(c). It looks like the result in Fig. 4.18(a) where the aliasing effect is noticeable. As expected, the high frequency information of the die is lost in the downsampling process. A bilinear interpolation is carried out on this downsampled hologram to have a new hologram with the resolution of 1024×1024. The reconstructed result of this interpolated hologram is

shown in Fig. 4.22(d). Comparing Fig. 4.22(d) with (b), the high frequency information is not restored by bilinear interpolation so that the image is still suffering from serious aliasing effects. Similarly, the hologram in Fig. 4.22(a) is bilinearly interpolated to a new hologram with resolution of 2048×2048 . Making a comparison of its reconstructed result in Fig. 4.22(e) and the result from a super hologram taken by 4-scan mode shown in Fig. 4.22(f), Fig. 4.22(f) provides better noise reduction and correct result. However in Fig. 4.22(e), unwanted replica of the twin image and the image of the die are obvious. Therefore the interpolation cannot restore the high frequency information lost or damaged by LR sampling but SR algorithms can achieve it.

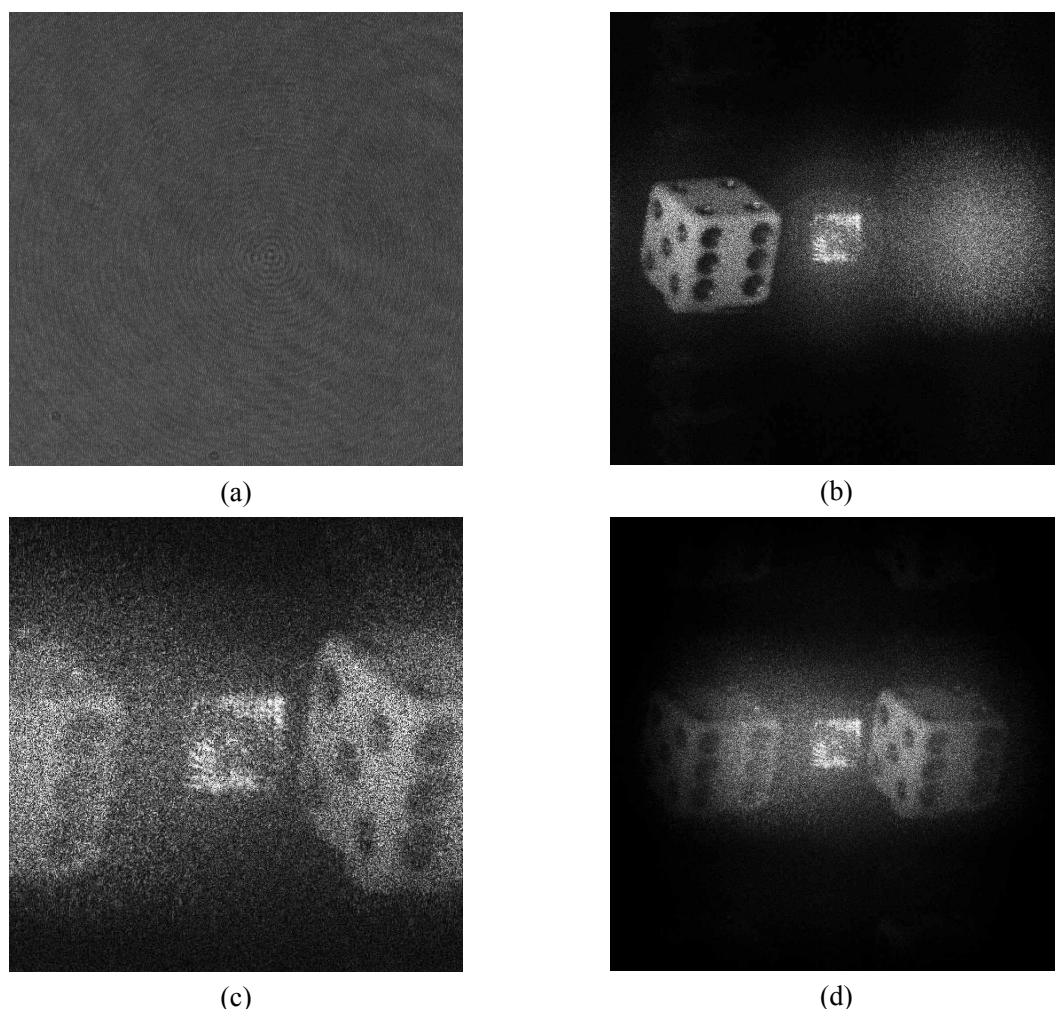


Fig. 4.22 (a) a hologram of a die with resolution of 1024×1024 ; amplitude-contrast images reconstructed from (b) hologram in (a); (c) a downsampled hologram of (a) with resolution of 512×512 ; (d) a hologram bilinearly interpolated to 1024×1024 from the downsampled hologram.

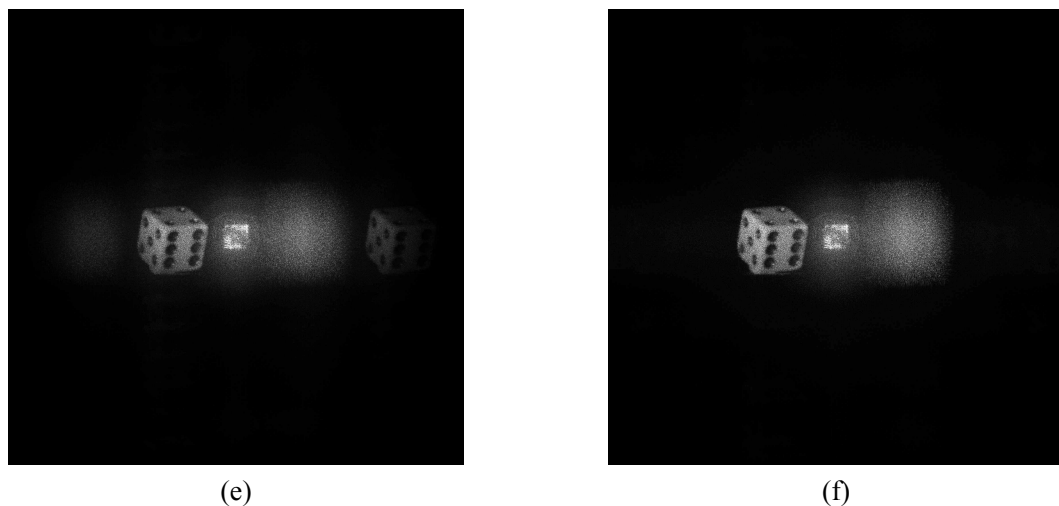


Fig. 4.22 (Continued) (e) a hologram bilinearly interpolated to 2048×2048 from (a); (f) a super hologram with resolution of 2048×2048 obtained from the camera by 4-scan mode.

As the experimental results above show, it has been verified that SR image reconstruction can be used in digital holography to increase the recording angle between the object wave and the reference wave. Compared to conventional methods of using translation stages to cause displacements of the whole camera, the ProgRes[®] MF^{scan} camera integrates a PZT in the body of the camera to control movement of the sensor chip alone at high precision. Therefore, the ProgRes[®] MF^{scan} camera and the camera's function of "microscanning" are very suitable to performing SR image reconstruction. The use of SR image reconstruction in phase application, such as holographic interferometry, will be demonstrated in Chapter 5.

4.5 Conclusions

In this chapter, various methods for improving the performance of digital holography have been discussed. The methods used for suppression of the zero-order term and the twin image were discussed previously in Chapter 3. Therefore here the discussion has been focused on three main methods for improving digital holography.

Firstly, the intensity ratio of the reference wave and the object wave was found to affect the quality of the reconstructed results in digital holography, just as it does in normal interferometric applications. Therefore it is important to keep the intensity of both waves comparable in order to obtain good results.

Secondly, as a coherent illumination source is used, speckle noise is unavoidable in digital holography due to the object's rough surface. Many approaches have been developed to reduce speckle noise so that better resolution and image quality can be achieved. In terms of ease and flexibility of manipulation, digital image processing methods seem to be the best choice to deal with speckle noise. However, the lack of a fully proven formula and the usual image processing dependence on "rule of thumb" is a disadvantage. Multiple-hologram methods are widely used to reduce speckle noise. The multiple holograms are obtained by varying the speckle pattern of the object wave. This can be done by changing the status of the wave used to illuminate the object via translation stages, etc. This increases the complexity of the experimental system, but has the advantage of increased spatial resolution and reducing levels of speckle noise. Deconvolution of the aperture function is a newly proposed method. As with the digital image processing methods, it does not need more than one hologram to be recorded. Thus the experimental system is simpler and the time required to capture and process the hologram is shorter than that of multiple-hologram methods. However considering the results shown in Fig. 4.11, the improvements brought about by this method are limited when compared with the other methods mentioned.

Thirdly, the low resolution of the CCD sensor results in heavy aliasing when it is applied to measure or visualize objects of relatively large size, e.g. of the order of tens of millimetres. It is also very hard to observe objects at closer distances due to the limited resolution of the CCD sensor. In our work, the super-resolution image reconstruction approach has been incorporated into our measuring system via a camera with microscanning capability, the ProgRes[®]MF^{scan} camera. This function not only removes aliasing to make the measurement more accurate, but also enables better quality reconstruction results to be obtained, with much larger observation ranges or alternatively a closer measurement distance for the same object. The use of the ProgRes[®]MF^{scan} camera enables the high accuracy and straightforward implementation of super-resolution image reconstruction algorithms.

To the author's best knowledge, this is the first time that holograms have been captured in super resolution by a camera with microscanning capability. The effect of super resolution approaches to digital holography has been thoroughly evaluated in this work. Also experimental work has been performed to propose a suitable approach to optimise the intensity ratio of the object beam and the reference beam.

References:

- BALDI, A. (2006) Comparative analysis of super-resolution algorithms for digital holography. Nimes, France, International Society for Optical Engineering, Bellingham WA, WA 98227-0010, United States.
- BAUMBACH, T., KOLENOVIC, E., KEBBEL, V. & JUPTNER, W. (2006) Improvement of accuracy in digital holography by use of multiple holograms. *Applied Optics*, 45, 6077-6085.
- BINET, R., COLINEAU, J. & LEHUREAU, J.-C. (2002) Short-range synthetic aperture imaging at 633 nm by digital holography. *Applied Optics*, 41, 4775-4782.
- CAI, X.-O. & WANG, H. (2008) The influence of hologram aperture on speckle noise in the reconstructed image of digital holography and its reduction. *Optics Communications*, 281, 232-7.
- DAINTY, J. C. (1975) *Laser Speckle and Related Phenomena*, New York, Springer-Verlag.
- DUBOIS, F., REQUENA, M. L. N., MINETTI, C., MONNOM, O. & ISTASSE, E. (2004) Partial spatial coherence effects in digital holographic microscopy with a laser source. *Applied Optics*, 43, 1131-1139.
- GARCIA-SUCERQUIA, J., RAMIREZ, J. A. H. & PRIETO, D. V. (2005) Reduction of speckle noise in digital holography by using digital image processing. *Optik*, 116, 44-48.
- GARCIA-SUCERQUIA, J., RAMIREZ, J. H. & CASTANEDA, R. (2006) Incoherent recovering of the spatial resolution in digital holography. *Optics Communications*, 260, 62-67.
- KANG, M. G. & CHAUDHURI, S. (2003) Super-resolution image reconstruction. *IEEE Signal Processing Magazine*, 20, 19-20.
- KORNIS, J. & GOMBKOTO, B. (2005) Application of super image methods in digital holography. Munich, Germany, International Society for Optical Engineering, Bellingham WA, WA 98227-0010, United States.
- KOZMA, A. & CHRISTENSEN, C. R. (1976) EFFECTS OF SPECKLE ON RESOLUTION. *Journal of the Optical Society of America*, 66, 1257-1260.
- KREIS, T. & SCHLUTER, K. (2007) Resolution enhancement by aperture synthesis in

- digital holography. *Optical Engineering*, 46, 055803.
- LIU, C., LIU, Z., BO, F., WANG, Y. & ZHU, J. (2002) Super-resolution digital holographic imaging method. *Applied Physics Letters*, 81, 3143.
- LOWENTHA, S. & JOYEUX, D. (1971) SPECKLE REMOVAL BY A SLOWLY MOVING DIFFUSER ASSOCIATED WITH A MOTIONLESS DIFFUSER. *Journal of the Optical Society of America*, 61, 847-&.
- MASSIG, J. H. (2002) Digital off-axis holography with a synthetic aperture. *Optics Letters*, 27, 2179 - 2181.
- PARK, S. C., PARK, M. K. & KANG, M. G. (2003) Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine*, 20, 21-36.
- QUAN, C., KANG, X. & TAY, C. J. (2007) Speckle noise reduction in digital holography by multiple holograms. *Optical Engineering*, 46, 115801.
- SCHNARS, U. & JUPTNER, W. (2005) *Digital holography - digital hologram recording, numerical reconstruction, and related techniques*, Springer.
- SCHNARS, U., KREIS, T. M. & JUPTNER, W. P. O. (1996) Digital recording and numerical reconstruction of holograms: reduction of the spatial frequency spectrum. *Optical Engineering*, 35, 977-82.

5 Optical contouring of 3D objects

5.1 Introduction

As illustrated in Chapter 1, shape measurement plays an important part in many applications in engineering and science. Examples of those applications include inspection and quality control in industry, robotics, biomedicine, the measurement of mechanical wear, and 3D object surface data recording for research into deformation and strain analysis. Optical methods for 3D shape measurement have been studied extensively because they are noncontact, nondestructive and are able to detect the whole measurement field and are therefore high speed. These optical contouring methods can be categorised into both incoherent illumination methods, such as fringe projection, moiré, and into coherent illumination methods, such as holographic interferometry, speckle interferometry, etc. Amongst these optical methods, interferometry offers 3D surface shape measurement at high resolution. Well-known interferometric methods for 3D shape measurement include holography and speckle interferometry. Quantitative evaluation of fringe patterns was first applied in Electronic Speckle Pattern Interferometry (ESPI) to process the fringes formed upon the object's surface. During the recording stage in ESPI, a lens is used to image the fringes and this introduces aberrations and makes the optical system more complex. However, no lens is used in holographic interferometry. The whole system is simpler and is therefore less influenced by aberrations. Conventional holographic interferometry performs well for many applications, as discussed previously in section 1.3. However in terms of the flexibility and processing speed it cannot meet the requirements of industrial environments. In this chapter, the feasibility of applying digital holography to the application of 3D surface contouring will be explored.

In digital holography, the reconstructed optical field from the digital hologram is calculated as a set of complex numbers, through which the amplitude- and phase-contrast images can be obtained. When imaging microscopic objects, phase-contrast images are of great interest as the 3D contour and absolute refractive indices of these objects can be determined from their phase-contrast images (Cuche et al., 1999). Phase-contrast imaging is used extensively and quantitatively in the microscopic field. However, it is of limited use with macroscopic objects. Macroscopic objects may be defined as objects with dimensions at

least 4 orders of magnitude greater than the wavelength of the optical field. Because their roughness is comparable to the wavelength of the optical wave, their phase-contrast images consist of values randomly distributed between the range $-\pi$ and $+\pi$. Therefore, a point-to-point subtraction of two phase-contrast images is necessary to determine the 3D contours of large objects. The result of this subtraction produces fringes which are absolute contours of surface height above some reference surface. Usually these two phase-contrast images of the same macroscopic object are obtained by using either two different illumination sources, two different wavelengths or two different refractive indices.

In this chapter, both conventional and digital implementations of these three methods, two-source, two-wavelength and two-refractive-index methods, will be discussed in detail. Finally a brief comparison of these three techniques is presented.

5.2 Contour generation

As shown in previous chapters, the image formed by the wavefront reconstruction process in holography is a replica of the object. This image contains a multiplicative phase function which is often absorbed into the image term and ignored (Hildebrand and Haines, 1967). However, this phase function is fundamental to optical contouring by holography. According to Fig. 5.1, this phase term is

$$\varphi = \mp k_1(r'_0 + r_1 - r_2) + k_2(r_a - r_b) \quad (5.1)$$

where $k_1 = 2\pi/\lambda_1$ and $k_2 = 2\pi/\lambda_2$, λ_1 is the wavelength used in the recording process and λ_2 is the wavelength used in the reconstruction process.

If contours are to be generated on the image, obviously additional information about the object must be obtained. If the source that is used to illuminate the object is changed in position during exposures, the hologram sees the object under illumination from two different angles. Likewise, multiple-frequency illumination provides additional information about the object because the complex reflectivity of the object is a function of frequency. Another way to interpret this is that the change of φ in Eq. (5.1) produces the desired interference pattern. Therefore, changes in the optical path, illumination wavelength or refractive index of the media results in the formation of contour fringes.

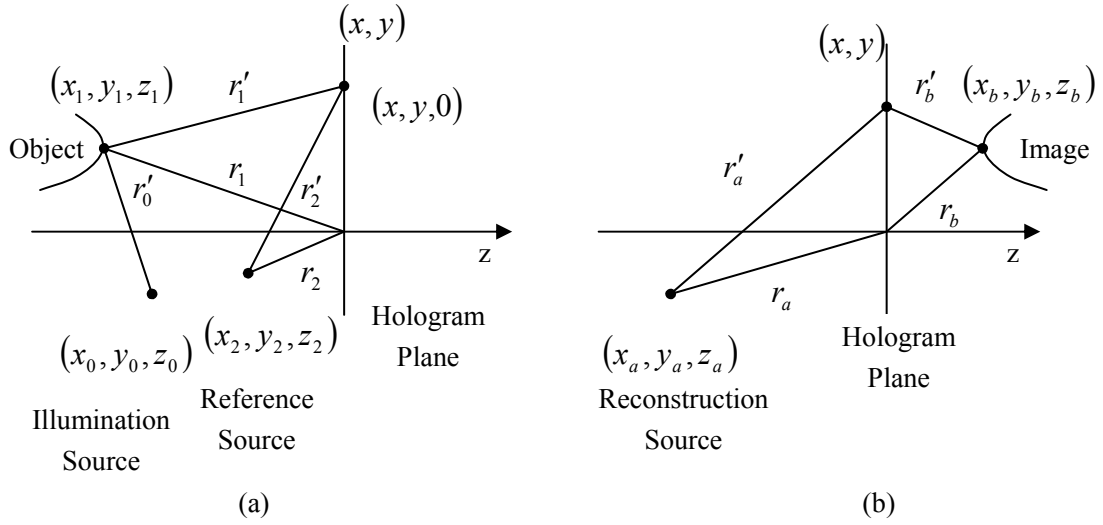


Fig. 5.1 Geometries for (a) the hologram recording process; (b) the reconstruction process.

5.3 Two-source methods

5.3.1 The conventional two-source method

In conventional holography, a double-exposure hologram may be formed on a holographic film by the two-source method. Between exposures the light source used to illuminate the object is shifted laterally by a small amount. Usually the object is illuminated by the reflection of the light source via an adjustable mirror. Through changing the angle of the mirror between exposures, the source appears to be translated. Hence, when this double-exposure hologram is reconstructed, a pattern of interference fringes is seen.

The principle of this method is quite straightforward. The two beams from two different light sources generate the planes of interference when they cross. The contours are formed when any object is placed into this light field, which appears to be cut by the planes of interference into a number of slices. In the recording process of the double-exposure hologram, the object is illuminated sequentially with each of the light sources, so no contours can be seen directly. But when this double-exposure hologram is reconstructed by the reference wave, both illumination beams from these two light sources appear simultaneously then the contours are observable as if illuminated by the two crossed beams. From this it can also be seen that for collimated illumination beams the fringe or contour spacing Δh is given by

$$\Delta h = \frac{\lambda}{2\sin(\theta/2)} \quad (5.2)$$

where θ is the angle between the two illumination beams and λ is the wavelength of the light sources. The reference planes for the contours will bisect the two beams.

5.3.2 The two-source method in digital holography

For the measurement of 3D surface shape using the two-source method in digital holography, a typical setup is shown in Fig. 5.2.

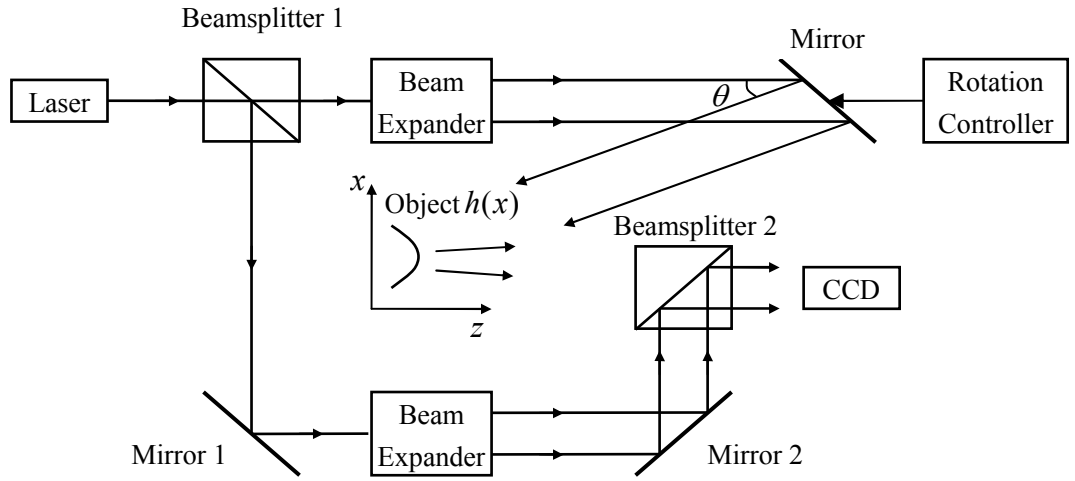


Fig. 5.2 A typical setup for surface contouring by the two-source method in digital holography

Beamsplitter 1 is used to separate the expanded laser beam into two beams: an object beam and a reference beam. The object beam illuminates the surface of the object via a rotatable mirror, whose movement is controlled by an electronic or a piezo-electric controller. The reference beam remains unchanged and is directed to the CCD sensor by mirror 1 and mirror 2. The reflected light from the surface of the object and the reference beam are combined by beamsplitter 2 and the interference pattern is captured by the CCD sensor.

By numerical reconstruction in digital holography, both amplitude- and phase-contrast images are obtained directly. Here only the phase-contrast images are of the interest for determining the 3D surface contours of macroscopic objects. Between the two exposures, the rotatable mirror is rotated by an angle $\Delta\theta/2$ from its initial angle of incidence θ . For simplicity, a two-dimensional analysis is adopted here. The object phase before and after the rotatable mirror tilt, $\varphi_{\theta_1}(x, -z_0)$ and $\varphi_{\theta_2}(x, -z_0)$, is expressed by

$$\varphi_{\theta_1}(x, -z_0) = k[x \sin \theta + h(x) \cos \theta + h(x)] \quad (5.3)$$

$$\varphi_{\theta_2}(x, -z_0) = k[x \sin(\theta + \Delta\theta) + h(x) \cos(\theta + \Delta\theta) + h(x)] \quad (5.4)$$

where x and $h(x)$ are shown in Fig. 5.2, $\Delta\theta$ is the change of the illumination angle between the two exposures, and k is the wave number. As previously mentioned the roughness of the surface is comparable with the optical wavelength so that each phase-contrast image is randomly distributed between the values $-\pi$ and π . Hence, the phase difference becomes

$$\begin{aligned} \Delta\varphi_{\theta}(x, -z_0) &\equiv \varphi_{\theta_2}(x, -z_0) - \varphi_{\theta_1}(x, -z_0) \\ &= 2k \sin \frac{\Delta\theta}{2} \left[x \cos\left(\theta + \frac{\Delta\theta}{2}\right) - h(x) \sin\left(\theta + \frac{\Delta\theta}{2}\right) \right] \end{aligned} \quad (5.5)$$

which means that the object surface is intersected by a group of planes parallel to the bisector of the illuminating beams and spaced by $\lambda/[2\sin(\Delta\theta/2)]$, where λ is the wavelength of the laser. The contour interval Δh is given by the following expression if the tilt angle is sufficiently small:

$$\Delta h = \frac{\lambda}{2 \sin \frac{\Delta\theta}{2} \sin\left(\theta + \frac{\Delta\theta}{2}\right)} \approx \frac{\lambda}{\Delta\theta \sin \theta} \quad (5.6)$$

It should be noted that Eq. (5.6) only holds for $x = 0$.

As the phase-contrast images φ_{θ_1} and φ_{θ_2} are calculated from the recorded holograms according to the algorithms mentioned in Chapter 2, they are modulo 2π phase images, i.e. wrapped phase images. The phase change induced by the translated illumination source monotonically increased or decreased within the object field, so the phase difference between these phase-contrast images $\Delta\varphi_{\theta}$ cannot be computed by simply subtracting image φ_{θ_1} directly from image φ_{θ_2} but instead is given by

$$\Delta\varphi_{\theta}(x, -z_0) = \begin{cases} \varphi_{\theta_2}(x, -z_0) - \varphi_{\theta_1}(x, -z_0), & \text{if } \varphi_{\theta_2}(x, -z_0) \geq \varphi_{\theta_1}(x, -z_0) \\ \varphi_{\theta_2}(x, -z_0) - \varphi_{\theta_1}(x, -z_0) + 2\pi, & \text{if } \varphi_{\theta_2}(x, -z_0) < \varphi_{\theta_1}(x, -z_0) \end{cases} \quad (5.7)$$

$\Delta\varphi_{\theta}(x, -z_0)$ is also a wrapped phase image. Conventional phase unwrapping algorithms may be applied to this phase difference image in order to remove the 2π ambiguities and so achieve an absolute phase difference map and the surface contours of the object.

To obtain contouring planes normal to the line of sight in reflective holography, the object must be illuminated from the side. The consequent problem of shadowing is unavoidable for the objects with complex surface shapes. Meanwhile, the first term in Eq. (5.5) is the linear phase introduced by the use of illumination from the side and this phase term is independent from the object. It must be removed in order to obtain the correct contours of the object related to the surface height distribution $h(x)$.

5.3.3 Experimental system for the two-source method

The scheme of our experimental system is similar to the geometry shown in Fig. 5.2, and Fig. 5.3 shows a photograph of the real system.

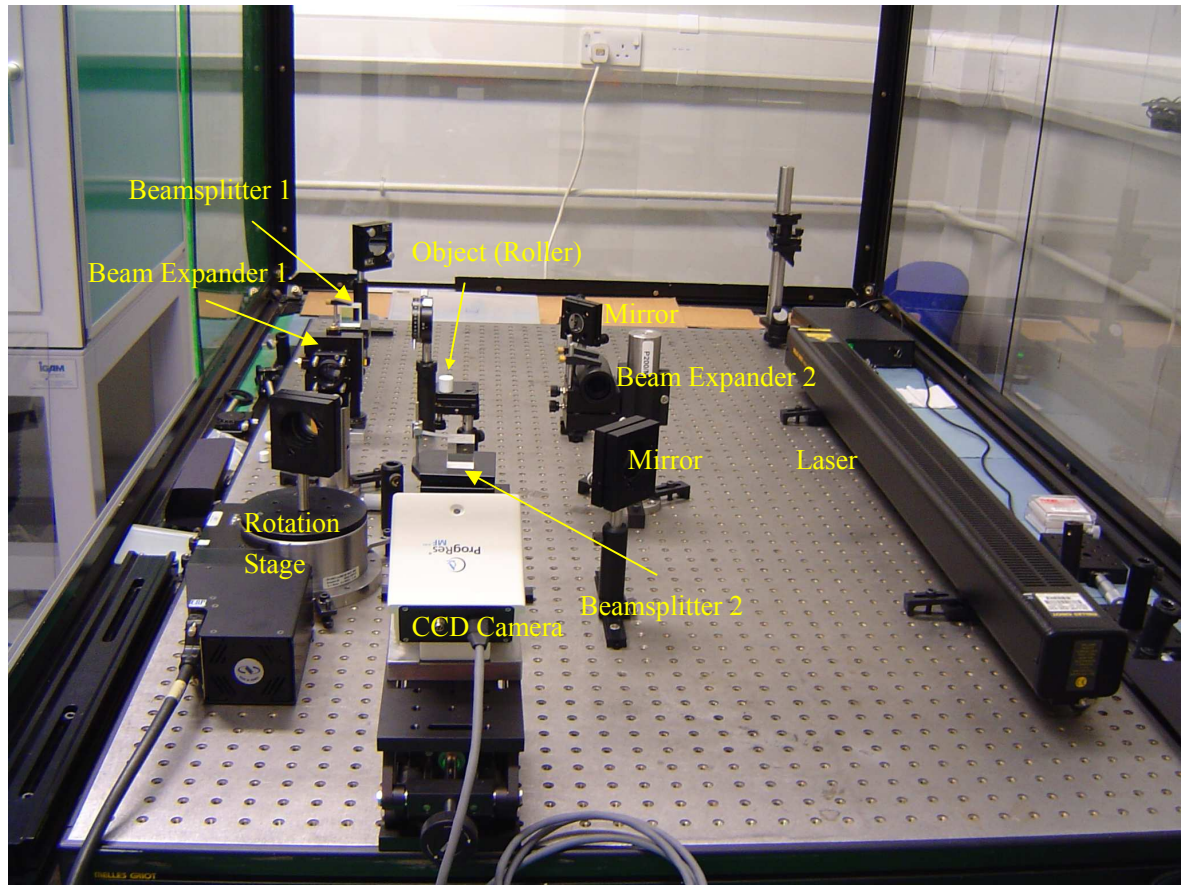


Fig. 5.3 A photograph of the experimental system

A 35mw HeNe laser (Melles Griot, model number of 25 LHP 928-230) with a wavelength of 632.8nm was used as the light source. The laser beam passes through a cubic beamsplitter of size 25.4mm×25.4mm×25.4mm to form two beams. One beam, which is the reference beam, travels through two mirrors and is expanded by a 20× beam expander

made by Melles Griot. Then it is reflected by another cubic beamsplitter so that it enters the camera. The other beam, i.e. the object beam, expanded by beam expander which consists of an objective ($NA = 0.62, f = 4mm$) and a collimation lens ($f = 75mm$), illuminates the surface of the object by way of a reflecting mirror which is mounted on a Newport rotation stage RV120CC. This rotation stage is controlled by a Newport controller, model number MM4005. The resolution of this rotation stage is 0.001° . The CCD camera used to capture the hologram was a ProgRes[®] MF^{scan} from Jenoptik. The resolution of the CCD sensor is 1360×1024 . But with built-in piezo-control mechanics in the camera, the CCD sensor can be shifted precisely by a fraction of the pixel spacing so that a much higher resolution can be achieved by combining these shifted images to form a super resolution image.

In a digital holographic system, the CCD sensor is the aperture of the system. Therefore, the numerical aperture of the system is related to the size of the CCD sensor and also the distance between the object and the CCD sensor. If the distance between the object and the CCD sensor is d , the pixel spacing of the CCD is Δx and the number of pixels in each dimension for the CCD is N , the numerical aperture of the system can be described as

$$NA = \frac{N\Delta x}{2d} \quad (5.8)$$

In an optical system, the numerical aperture characterizes the range of angles over which the system can accept or emit light. The focal depth of the reconstructed image is given by

$$FL = \frac{\lambda}{(NA)^2} \quad (5.9)$$

Therefore, to measure an object at a distance of a few hundred millimetres, say $700mm$, with a recording wavelength of $632.8nm$ and a CCD sensor side length of $6mm$, the focal depth of the reconstructed image would be $34mm$. However for a distance of $200mm$, under the same conditions, the focal depth would be $2.8mm$. This shows that the object distance does not have to be measured precisely if the object is placed at a relatively large distance from the CCD. When the object is placed near the CCD, especially in microscopic applications, the distance should be measured precisely. However, the numerical focusing ability of digital holography can extend the focal depth in these applications to compensate for the shortages inherent in conventional microscopy.

5.3.4 Experimental results for the two-source method

Firstly a white-painted stainless steel roller bearing element was used as the measurement object. The radius of the roller was 5.60mm and its height is 11.20mm . Fig. 5.4 shows a photograph of this roller.



Fig. 5.4 A photograph of the roller under study

The roller was illuminated by the HeNe laser from an angle θ with respect to the optical axis of the CCD sensor, i.e. z axis. It is placed 527mm away from the CCD plane. The angle θ was set to -36° and the mirror was rotated with $\Delta\theta = 0.03^\circ$ between the two exposures. The phase-contrast images produced directly from the numerically reconstructed holograms for these two angles are shown in Fig. 5.5. Fig. 5.5(a) is the phase-contrast image with an illumination angle $\theta = -36.00^\circ$ and Fig. 5.5(b) is the phase-contrast image with an illumination angle $\theta = -35.97^\circ$. Both images look qualitatively the same. The phase of the reconstructed object field is randomly dispersed in the range of $-\pi$ to $+\pi$. As illustrated in section 5.1, these phase-contrast images do not offer direct phase information of the object's surface because of the surface roughness which is much higher than the wavelength of the illumination source. But the phase-difference image can provide the wrapped phase map of the object under study. Fig. 5.5(c) shows the phase-difference image which is obtained by subtracting Fig. 5.5(b) from Fig. 5.5(a) according to Eq. (5.7). Because the object is illuminated from the side, as shown in Fig. 5.3, there is a linear carrier term that must be deducted. Otherwise the correct object surface will not be achieved after phase unwrapping. Fig. 5.5(d) is the wrapped phase map of the object under study after removing the linear carrier term in Eq. (5.5). Fig. 5.5(e) is the unwrapped phase map of the roller. Fig. 5.5(f) is the reconstructed 3D surface of the roller.

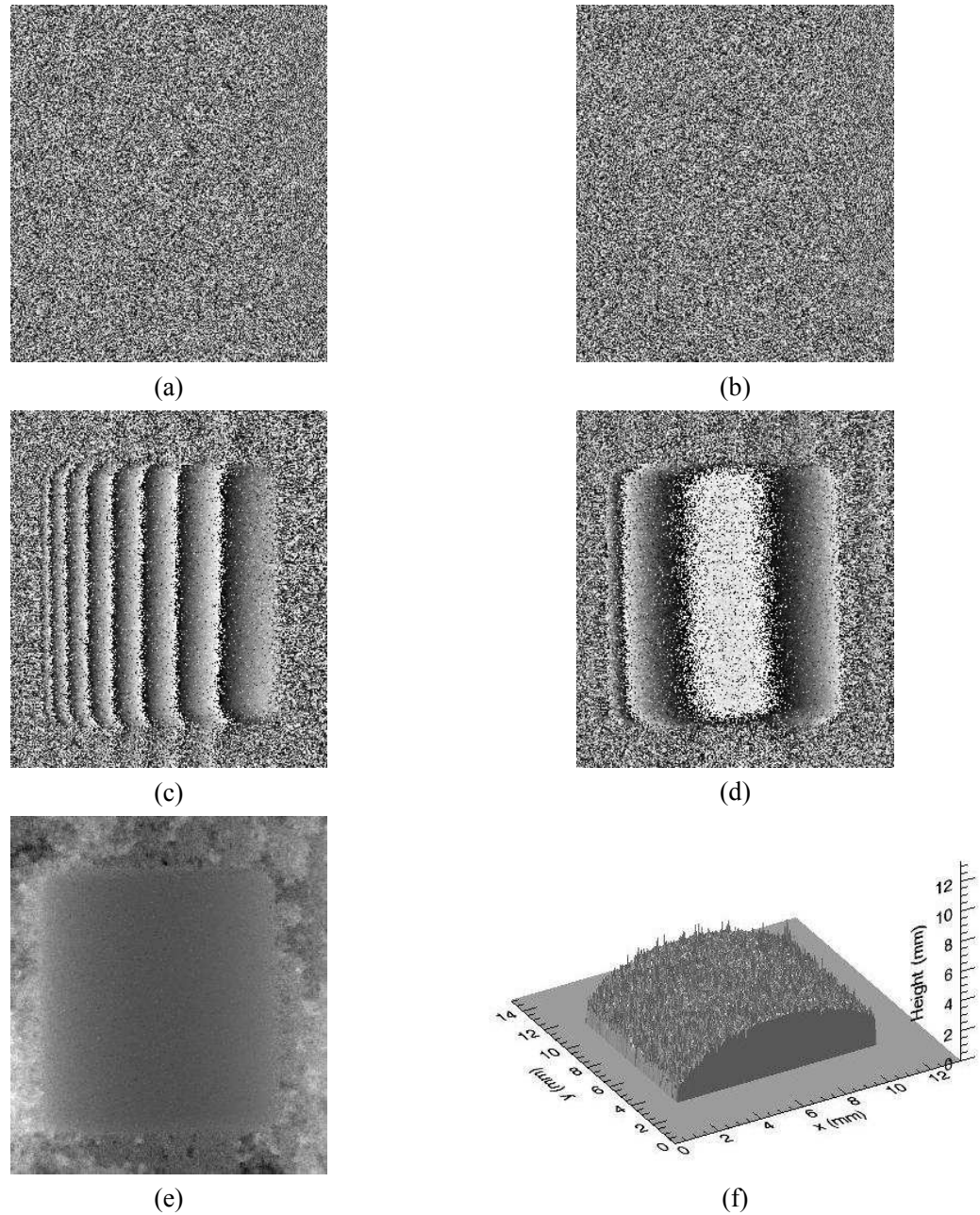


Fig. 5.5 Experimental results for a roller: the reconstructed phase-contrast images when (a) the illumination angle $\theta = -36^\circ$; (b) the illumination angle $\theta = -35.97^\circ$; (c) phase-difference image obtained by subtracting (b) from (a); (d) phase-difference image with the removal of the linear carrier term; (e) unwrapped phase-difference image; (f) the 3D height distribution of the object surface.

In the whole process of calculating the 3D height distribution of the object, there are some important aspects to note:

A. the removal of the linear carrier term

This is very important for the two-source method. The linear carrier factor has to be removed as tilted illumination is always used in this method. Otherwise the unwrapped

phase cannot provide the correct phase information. According to Eq. (5.5), the linear carrier factor t_l is given by:

$$t_l = 2kx \sin(\Delta\theta/2) \cos(\theta + \Delta\theta/2) \quad (5.10)$$

The calculated phase-difference term $\Delta\varphi_\theta$ in Eq. (5.7) is wrapped between 0 and $+2\pi$. However, t_l is beyond this range. To remove the linear carrier factor effectively, t_l has to be wrapped between 0 and $+2\pi$ as well. The arctan function is applied to Eq. (5.10) to obtain the wrapped linear carrier factor t_{lw} which is distributed between the range of 0 and $+2\pi$. The removal of the linear carrier factor is then calculated by

$$\varphi_{nw}(x, -z_0) = \begin{cases} t_{lw}(x, -z_0) - \Delta\varphi_\theta(x, -z_0), & \text{if } t_{lw}(x, -z_0) \geq \Delta\varphi_\theta(x, -z_0) \\ t_{lw}(x, -z_0) - \Delta\varphi_\theta(x, -z_0) + 2\pi, & \text{if } t_{lw}(x, -z_0) < \Delta\varphi_\theta(x, -z_0) \end{cases} \quad (5.11)$$

This new phase term φ_{nw} represents the actual phase introduced in the optical field by the surface shape of the roller. After inverting this term, a 3D representation of the surface of the roller can be obtained. Fig. 5.5(c) and (d) show the phase-difference image before and after the removal of the linear carrier factor.

B. the noise reduction of the wrapped phase map by digital image processing

As illustrated in previous chapters, because of the limited size of the CCD sensor, a fundamental problem in digital holography is the presence of speckle noise in the reconstruction process. The speckle pattern can be clearly seen in Fig. 5.5. The direct measurement result in Fig. 5.5(f) is noisy and cannot show the shape of the surface clearly. Besides the speckle noise, electronic noise from the CCD sensor also noticeably degrades the holograms, because the light reflected from the surface of the object is very weak. Also in such noisy wrapped phase-difference images, the phase unwrapping algorithm may detect false 2π jumps, which may cause additional image corruption effects. To avoid these problems, digital image processing techniques may be used to enhance the wrapped phase-difference image.

Median filters are commonly used to smooth the phase-difference images (Schnars, 1994). Median filtering can be applied to all the pixels in the phase-contrast image using a small window of variable size. Usually a 3×3 window is used. The method takes all the data points contained within the window, sorts them until the median is found, and then places

the median value into a location in another array corresponding to the centre of the window. The window is then moved across the image by one pixel, and the procedure is repeated until the entire array has been processed. Since the median filter does not create new pixel values when the filter encounters an edge, it is much better at preserving sharp edges than mean averaging filters. But phase itself is not a signal. It is instead a property of a signal. It is usually suggested that it is the observations which generate the phase rather than phase itself that must be filtered. However in digital holography, the original holograms are not available to be filtered and so the phase data itself must be filtered. Therefore the phase data cannot be filtered in a conventional way.

Suppose the phase in the phase-difference image is $\varphi(x, y)$. In the conventional way, the median filtered phase $\tilde{\varphi}(x, y)$ will be obtained by

$$\tilde{\varphi}(x, y) = \text{median}[\varphi(x, y)] \quad (5.12)$$

In fact, the phase data first must be mapped to vectors $\exp[j\varphi(x, y)]$ in the complex plane (Ghiglia and Pritt, 1998). These vectors can then be median filtered and the resulting phase extracted by means of the arc tangent operator. Because there is no natural way to define the median filter because of the lack of natural ordering for complex numbers, an approximation has to be defined by the equation

$$\tilde{\varphi}(x, y) = \arctan\{\text{median}[\sin \varphi(x, y)], \text{median}[\cos \varphi(x, y)]\} \quad (5.13)$$

where the median is evaluated in the 3×3 neighbourhood of the pixel (x, y) . Some researchers use other filters, such as the anisotropic sine/cosine average filter, to enhance the wrapped phase and prevent image corruption effects (Palacios et al., 2004). Here a median filter with a 3×3 window has been applied to the wrapped phase-difference image without the linear carrier factor (see Fig. 5.5(d)). Fig. 5.6(a) and (b) show the results of the wrapped phase-difference image filtered by Eq. (5.12) and Eq. (5.13), respectively. Fig. 5.6(c) and (d) show the unwrapped phase of Fig. 5.6(a) and (b). Fig. 5.6(e) and (f) are the 3D height distributions of the surface converted from the phase in Fig. 5.6(c) and (d). Fig. 5.6(g) shows the 3D height distribution of the surface in Fig. 5.6(f) by further smoothing of 3×3 window.

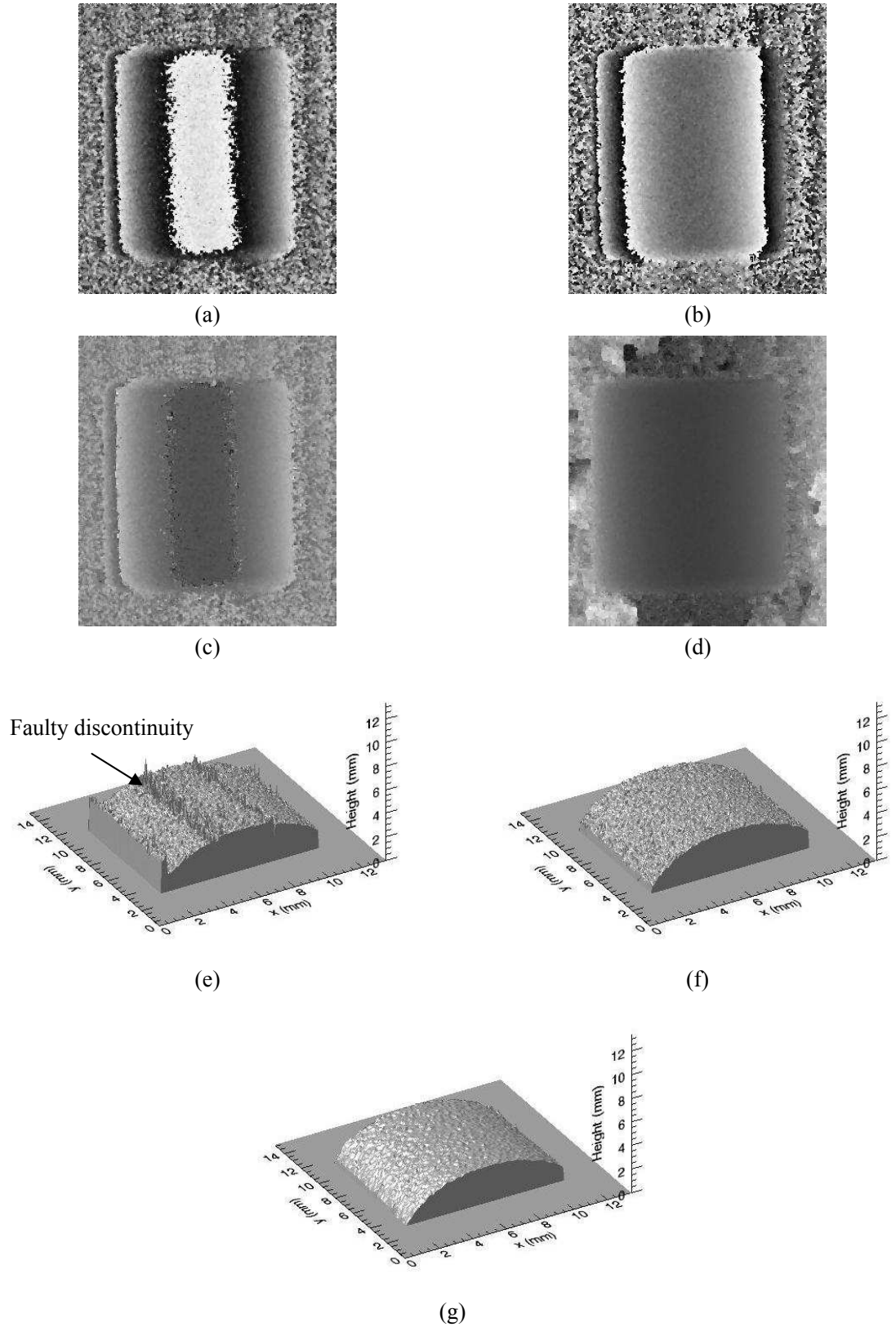


Fig. 5.6 (a) The wrapped phase image with incorrect median filtering in Eq. (5.12); (b) the wrapped phase image with correct median filtering in Eq. (5.13); (c) the unwrapped phase image from (a); (d) the unwrapped phase image from (b); (e) the 3D height distribution of the surface from (c); (f) the 3D height distribution of the surface from (d); (g) the 3D height distribution of the surface in (f) by further smoothing of 3×3 window.

Clearly from Fig. 5.6(e), a 2π jump has not been detected by the unwrapping algorithm and so there are discontinuities in the 3D representation of the surface. But when the wrapped phase-difference image is filtered correctly according to Eq. (5.13), the reconstructed surface shape (see Fig. 5.6(f)) can be seen to coincide with that of the roller under study. Though the application of median filter helps to reduce the speckle noise significantly, comparing Fig. 5.5(e) and Fig. 5.6(f), the reconstructed surface can be further smoothed by a 3×3 averaging window, as is shown in Fig. 5.6(g). By using a curve fitting technique, based on the Gradient-expansion algorithm (Bevington and Robinson, 1992), which combines the best features of the gradient search technique with the method of linearizing the fitting function, the measured radius of the roller in Fig. 5.6(g) was determined as $5.54mm$, which represents an error of only 1.1% compared with the value of $5.60mm$ as measured by a vernier calliper with an uncertainty of $0.02mm$.

Another nonlinear filtering method was proposed by Yamaguchi et al. (2001). The phase image is split into discrete 2×2 matrices and for each 2×2 sub-matrix a phase value for the pixel with the maximum amplitude of $|U_1 U_2^*|$ is chosen in order to compress the original data set. For an original phase-difference image of 1024×1024 pixels, the compressed image has a size of 512×512 pixels. This data compression is justified because the phase values at lower intensity pixels are less reliable. The results from this processing are presented in Fig. 5.7. Fig. 5.7(a) is the phase-difference image processed by this nonlinear filtering method, after removal of the linear inclination factor. Fig. 5.7(b) is the unwrapped phase map of Fig. 5.7(a). Fig. 5.7(c) is the direct 3D height distribution of the detected surface from Fig. 5.7(b). Fig. 5.7(d) is the result of further smoothing of Fig. 5.7(c) by a 3×3 window. Although there are some wrinkles on the surface of the roller, the shape is correctly obtained. Meanwhile the measured radius of the surface of $5.74mm$ has an error of 2.5% compared to that measured independently using a vernier calliper as $5.60\pm 0.01mm$.



Fig. 5.7 (a) The wrapped phase-difference image processed by this nonlinear filtering method ($\Delta\theta=0.03^\circ$); (b) the unwrapped phase image of (a).

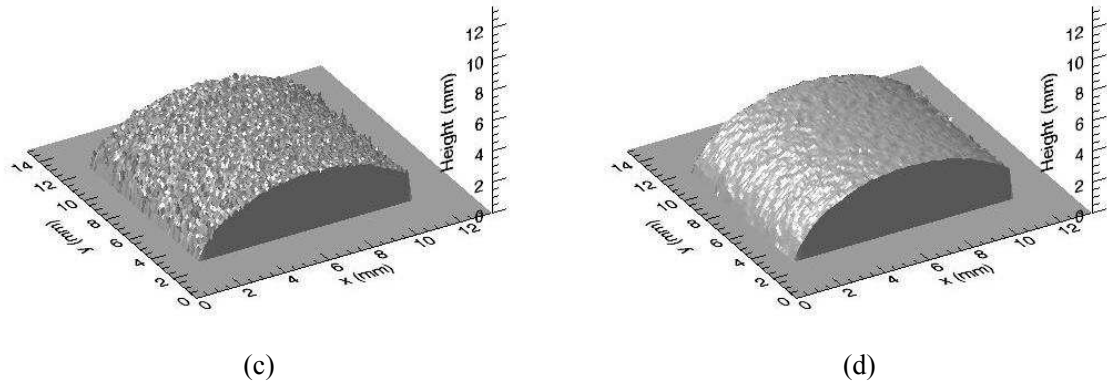


Fig. 5.7 (Continued) (c) the direct 3D height distribution of the measured surface from (b); (d) the further smoothing of (c) by a 3×3 window.

According to the principles of the filtering methods shown above, the nonlinear filtering method proposed by Yamaguchi et al. (2001) has an improved ability in suppressing speckle noise in digital holographic interferometry where the reflected light intensity from the surface of the object is often lower than would ideally be the case.

C. the effect of the application of different phase unwrapping algorithms

There are many different phase unwrapping algorithms that can be used for applications in interferometry. They can be classified into two categories: path-following methods and minimum-norm methods. Interested readers can find more details about phase unwrapping algorithms in the bibliography (Ghiglia and Pritt, 1998). Path-following methods are mainly based on branch-cut, quality-map or discontinuity techniques. They explicitly or implicitly generate branch cuts or lines of discontinuity. Goldstein's branch cut algorithm, the quality-guided path following algorithm, Flynn's minimum discontinuity algorithm and the recursive nearest neighbour algorithm all fall into among this category. Minimum-norm methods are completely different from the path-following methods. They impose constraints on the solution in a mathematically formal manner, in order to seek the unwrapped phase whose local derivatives match the measured derivatives as closely as possible, according to the generalized error norm. The minimum L^p -Norm algorithm and the PCG (preconditioned conjugate gradient) algorithm are typical methods in this category. Usually Goldstein's algorithm is the fastest amongst the phase unwrapping algorithms mentioned above. Its unwrapping results are good if the wrapped phase map is not noisy and is without discontinuities. In terms of success and failure rate, Flynn's algorithm and the minimum L^p -norm algorithm are better than the other algorithms. However, they run

considerably slower than Goldstein's algorithm. Fig. 5.8 shows results comparing different phase unwrapping algorithms. The wrapped phase-difference image in Fig. 5.8(a) is used to test all of these different phase unwrapping algorithms. It is the same as that shown previously in Fig. 5.5(d). The unwrapped phase maps and their corresponding 3D height distributions are displayed for comparison.

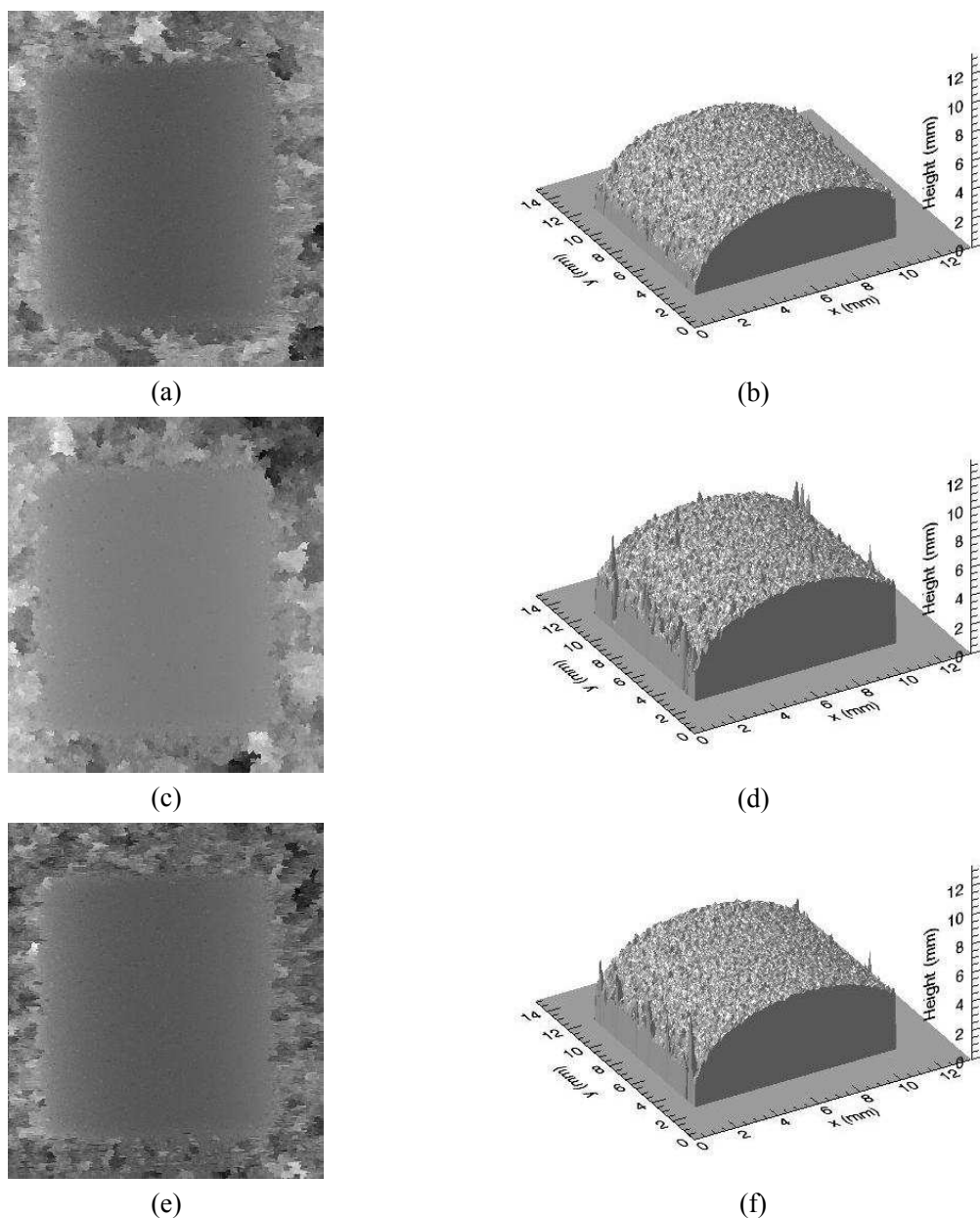
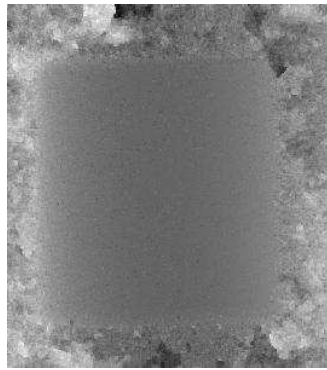
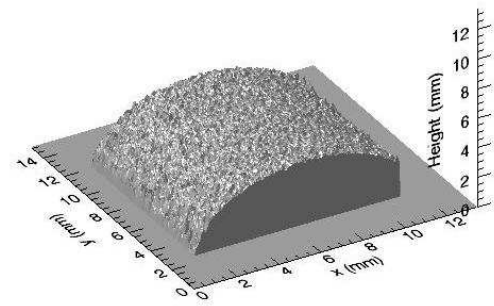


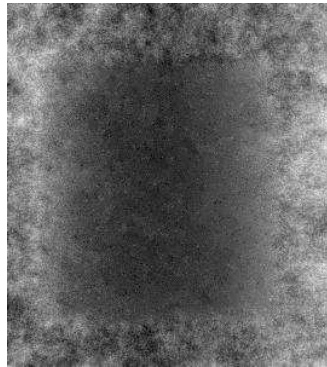
Fig. 5.8 A comparison of different phase unwrapping algorithms for the phase-difference image shown in Fig. 5.5(d): (a) the unwrapped phase map and (b) the height distribution produced using Goldstein's branch cut algorithm; (c) the unwrapped phase map and (d) the height distribution produced using the quality-guided path following algorithm; (e) the unwrapped phase map and (f) the height distribution produced using the mask cut algorithm.



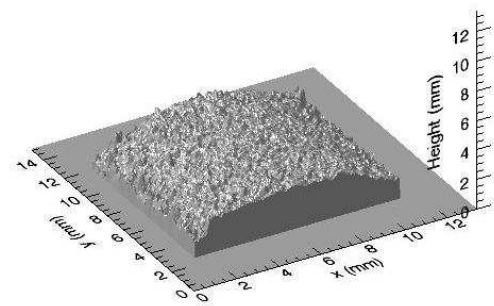
(g)



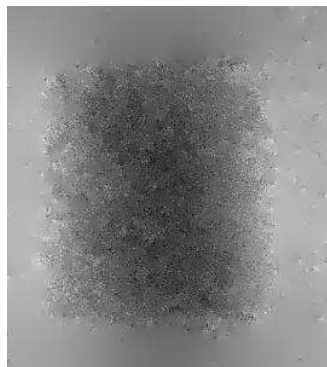
(h)



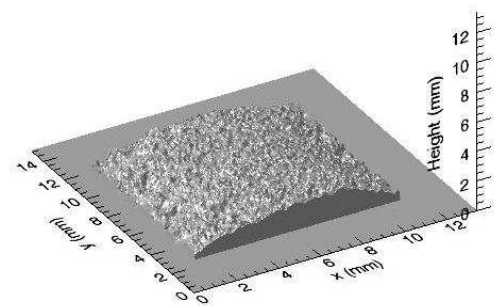
(i)



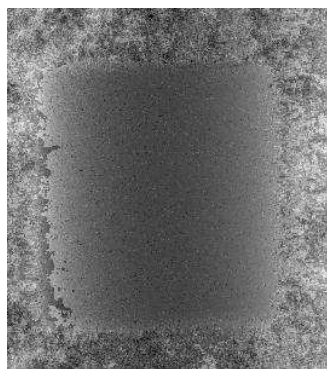
(j)



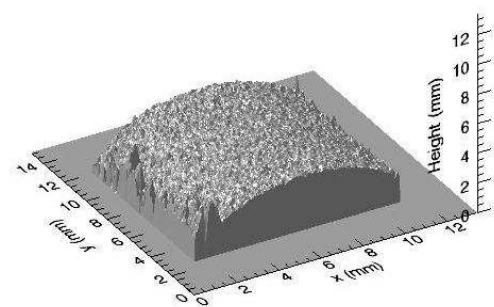
(k)



(l)

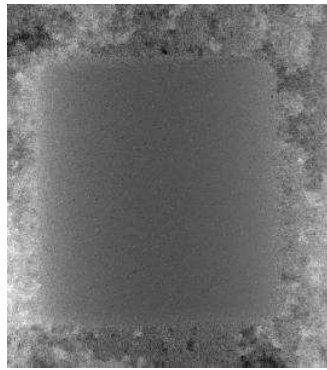


(m)

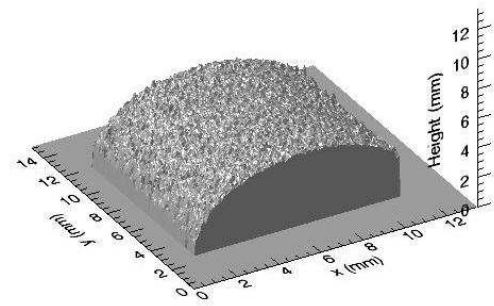


(n)

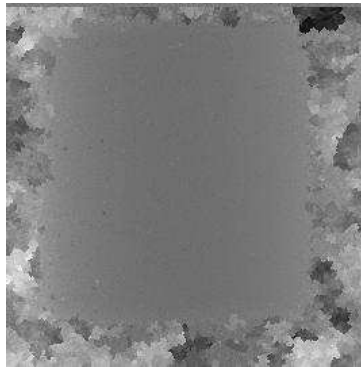
Fig. 5.8 (Continued) (g) the unwrapped phase map and (h) the height distribution produced using Flynn's minimum discontinuity algorithm; (i) the unwrapped phase map and (j) the height distribution produced using the Unweighted Least-squares (Unweighted Multigrid) algorithm; (k) the unwrapped phase map and (l) the height distribution produced using the PCG algorithm; (m) the unwrapped phase map and (n) the height distribution produced using the weighted multigrid algorithm.



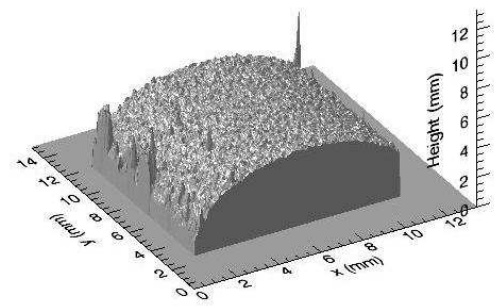
(o)



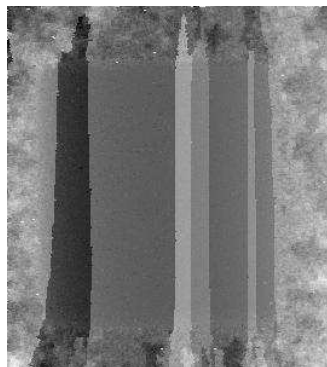
(p)



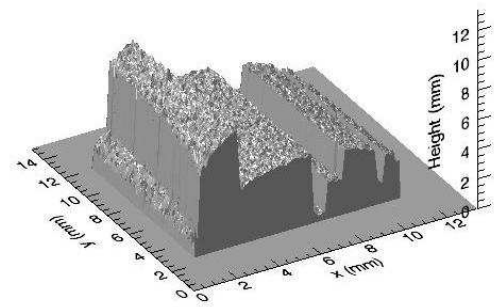
(q)



(r)



(s)



(t)

Fig. 5.8 (Continued) (o) the unwrapped phase map and (p) the height distribution produced using the minimum L^p -norm algorithm; (q) the unwrapped phase map and (r) the height distribution produced using the reliability ordering (Arevalillo-Herráez's) algorithm; (s) the unwrapped phase map and (t) the height distribution produced using the recursive nearest neighbor algorithm.

Phase Unwrapping Algorithm	Measured Radius of The Surface (mm)	Error (%)
Goldstein's Branch Cut Algorithm	5.538	-1.1
Quality-Guided Path Following Algorithm	5.686	1.5
Mask Cut Algorithm	5.646	0.8
Flynn's Minimum Discontinuity Algorithm	5.542	-1.0
Unweighted Least-Squares Algorithm	6.739	20.3
PCG Algorithm	8.760	56.4
Weighted Multigrid Algorithm	6.449	15.2
L^p -Norm Algorithm	5.569	-0.6
Reliability Ordering (Arevalillo-Herráez's) Algorithm	5.925	5.8
Recursive Nearest Neighbour Algorithm	N/A	N/A

Table 5.1 The measured radius of the roller according to Fig. 5.5(d)

Phase Unwrapping Algorithm	Measured Radius of The Surface (mm)	Error (%)
Goldstein's Branch Cut Algorithm	5.588	-0.2
Quality-Guided Path Following Algorithm	5.588	-0.2
Mask Cut Algorithm	5.588	-0.2
Flynn's Minimum Discontinuity Algorithm	5.588	-0.2
Unweighted Least-Squares Algorithm	6.190	10.5
PCG Algorithm	6.501	16.1
Weighted Multigrid Algorithm	5.822	4.0
L^p -Norm Algorithm	5.588	-0.2
Reliability Ordering (Arevalillo-Herráez's) Algorithm	5.588	-0.2
Recursive Nearest Neighbour Algorithm	N/A	N/A

Table 5.2 The measured radius of the roller according to Fig. 5.6(b)

Phase Unwrapping Algorithm	Measured Radius of The Surface (mm)	Error (%)
Goldstein's Branch Cut Algorithm	5.641	0.7
Quality-Guided Path Following Algorithm	5.603	0.1
Mask Cut Algorithm	5.641	0.7
Flynn's Minimum Discontinuity Algorithm	5.641	0.7
Unweighted Least-Squares Algorithm	6.466	15.5
PCG Algorithm	6.426	14.8
Weighted Multigrid Algorithm	5.641	0.7
L^p -Norm Algorithm	5.660	1.1
Reliability Ordering (Arevalillo-Herráez's) Algorithm	5.630	0.5
Recursive Nearest Neighbour Algorithm	N/A	N/A

Table 5.3 The measured radius of the roller according to Fig. 5.7(a)

As the results are shown above, different phase unwrapping algorithms have different performance. For the phase-difference images both with and without digital filtering shown in Fig. 5.5(d), Fig. 5.6(b) and Fig. 5.7(a), Goldstein's branch cut algorithm, quality-guided path following algorithm, mask cut algorithm, Flynn's algorithm and the L^p -norm algorithm performed stably with errors between -1.1% to 1.5%. However the recursive nearest neighbour algorithm failed to extract the correct 3D height distribution of the roller under any case. Clear discontinuities can be seen in Fig. 5.8(s), which indicated the failure of the algorithm to detect the 2π jumps. The digital filtering of the phase-difference image improved the results of the phase unwrapping process. Comparing Table 5.2 and Table 5.3 with Table 5.1, the error decreased after performing digital filtering on the phase-difference images. For the phase unwrapping algorithms which are quite sensitive to the noise, such as unweighted least-squares algorithm, PCG algorithm, weighted multigrid algorithm and reliability ordering algorithm, their performances improved significantly after the phase-difference images were digitally filtered. In terms of running time, all the algorithms took a similar time, in the range to two seconds, to process an image with a size of 266×300 pixels, except L^p -norm algorithm which took 15 seconds to process the same image.

D. the effect of the contour interval

The contour interval, defined by Eq. (5.6), is in inverse proportion to the number of contour fringes in optical contouring. To investigate its effect on the reconstructed 3D surface height distribution, various contour intervals were used. In the two-source contouring method, the height sensitivity is related to the illumination angles in both exposures and the recording wavelength according to Eq. (5.6). Fig. 5.10 to Fig. 5.14 show the results obtained by varying $\Delta\theta$ to achieve different height sensitivities.

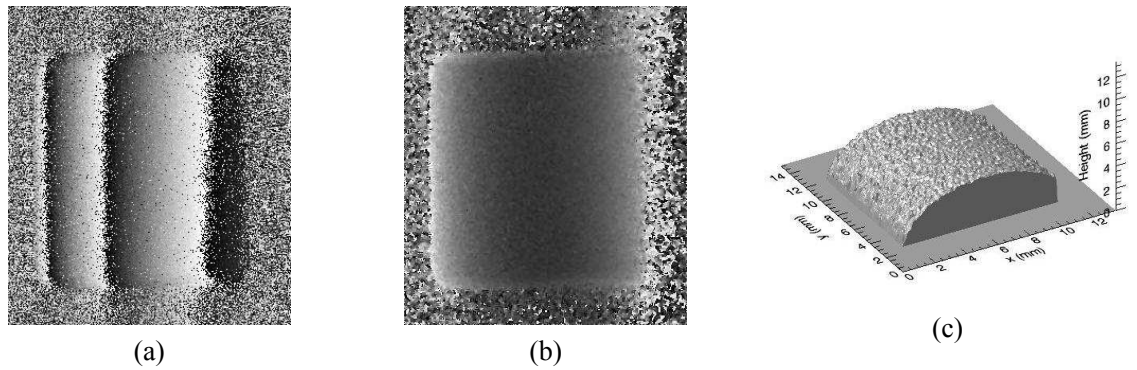


Fig. 5.9 The contouring results when $\Delta\theta=0.01^\circ$: the phase-difference images (a) with and (b) without illumination factor; (c) the 3D height distribution of the roller.

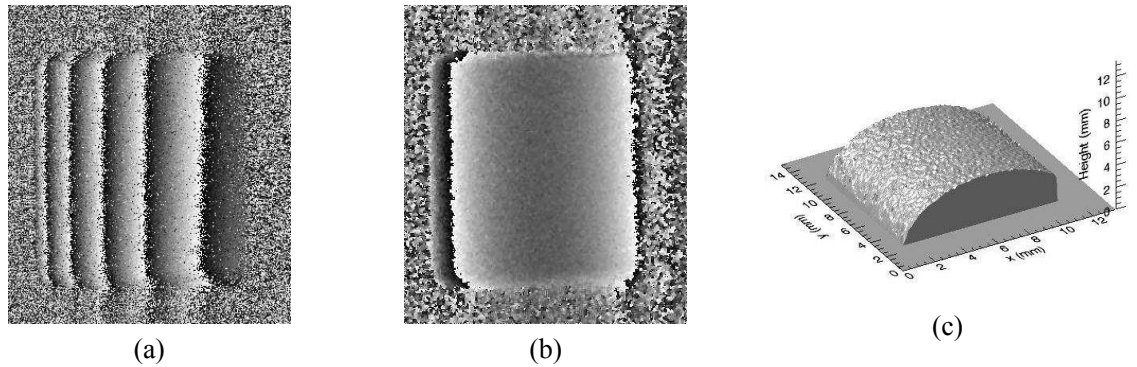


Fig. 5.10 The contouring results when $\Delta\theta=0.02^\circ$: the phase-difference images (a) with and (b) without illumination factor; (c) the 3D height distribution of the roller.

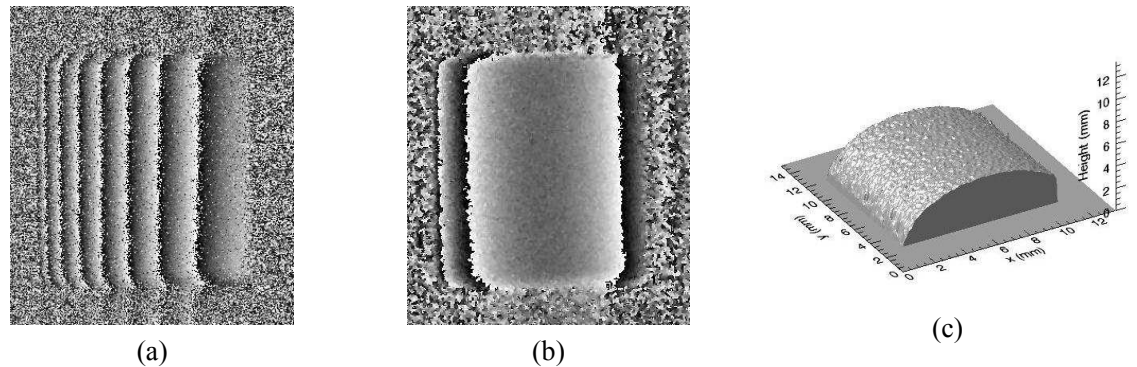


Fig. 5.11 The contouring results when $\Delta\theta=0.03^\circ$: the phase-difference images (a) with and (b) without illumination factor; (c) the 3D height distribution of the roller.

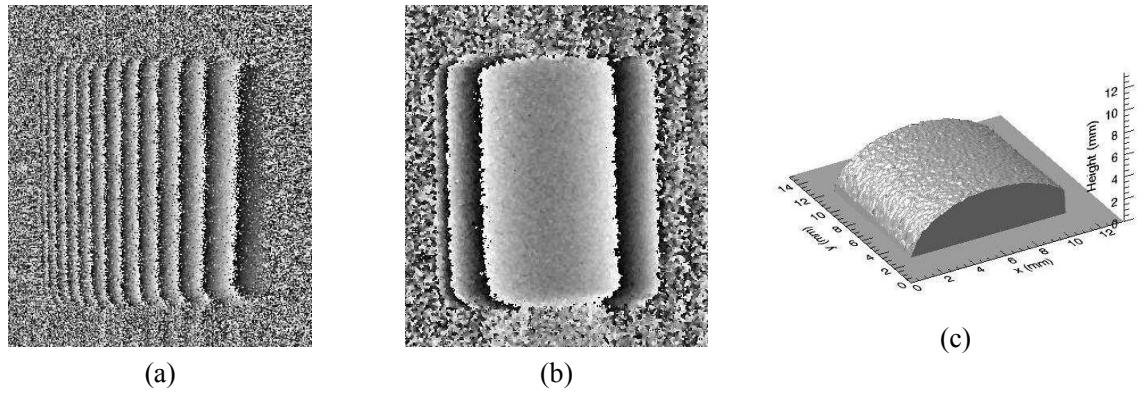


Fig. 5.12 The contouring results when $\Delta\theta=0.04^\circ$: the phase-difference images (a) with and (b) without illumination factor; (c) the 3D height distribution of the roller.

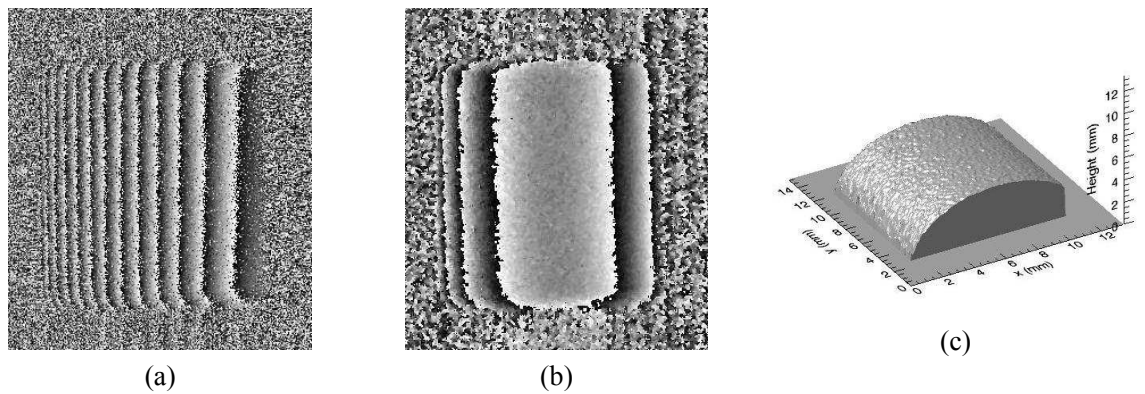


Fig. 5.13 The contouring results when $\Delta\theta=0.05^\circ$: the phase-difference images (a) with and (b) without illumination factor; (c) the 3D height distribution of the roller.

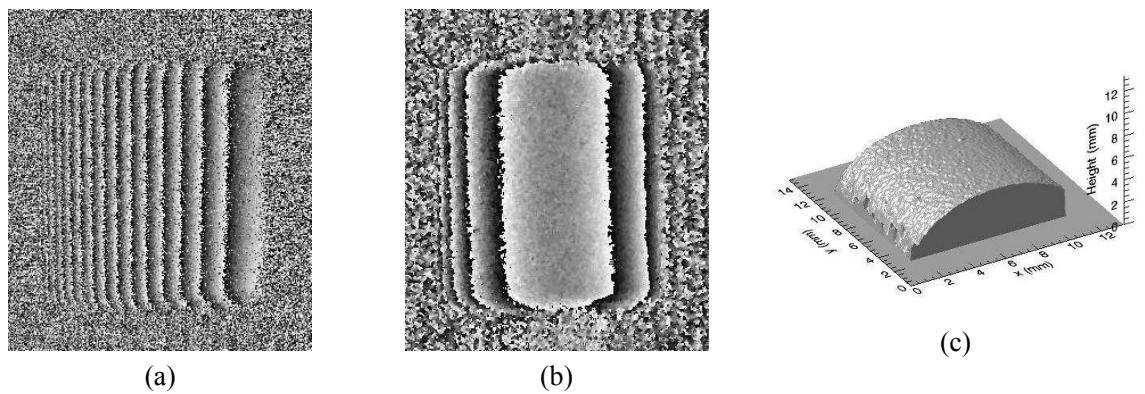


Fig. 5.14 The contouring results when $\Delta\theta=0.06^\circ$: the phase-difference images (a) with and (b) without illumination factor; (c) the 3D height distribution of the roller.

Contouring Results	Illumination Angle Change ($^{\circ}$)	Contour Interval (mm^{-1})	Measured Radius of the Roller (mm)	Error (%)
Fig. 5.9	0.01	6.17	5.63	0.5
Fig. 5.10	0.02	3.08	5.64	0.7
Fig. 5.11	0.03	2.06	5.59	0.2
Fig. 5.12	0.04	1.54	5.62	0.4
Fig. 5.13	0.05	1.23	5.60	0.0
Fig. 5.14	0.06	1.03	5.65	0.9

Table 5.4 The numerical results of changing contour interval by altering $\Delta\theta$

From the figures and numerical results shown above, the variation of contour interval via altering the change of illumination angle did not make a significant difference to the accuracy of the measured radius. The measurement error remained in the range of 1% for all different contour intervals tested. The number of contours increases when the contour interval decreases. For large contour interval, there is no contour in the phase-difference image when the illumination factor is removed, as shown in Fig. 5.9, which means no phase unwrapping algorithm is needed in order to extract the 3D representation of the roller. As for the phase-difference image with contours displayed, a phase unwrapping algorithm is necessary. In this example, mask cut algorithm was used to obtain the 3D representation of the roller.

Theoretically a decrease in contour interval improves the system's ability to measure fine surface detail. However too small a contour interval will not perform well in holographic interferometry. Limited by the small size of the CCD sensor and the weak intensity reflected from the object, noise destroys the quality of the finer contours. Meanwhile, the low resolution of the CCD sensor cannot resolve the contours denser than the pixel spacing of the camera. This is shown in Fig. 5.14. The contrast of the contours in this figure is less than it in Fig. 5.13 and the edge of the roller was not recovered successfully. Therefore, the contour interval has to be chosen carefully according to the specific conditions dictated by the object and the optical system.

The comparisons shown in Table 5.1 to 5.4 led us to the idea of using the two-source contouring technique as a numerical method for evaluating the effects caused by the various different processing techniques. In the previous chapters, the choices for finding

suitable processing methods, such as for the suppression of the zero-order term and for the reduction of speckle noise, etc., are based purely on the visual quality of the reconstructed images. This is a subjective process. However through the use of the two-source contouring technique to quantitatively measure an object of accurately known dimensions by varying the processing methods in due course, these quantitative results can be used as a quantitative benchmark in order to evaluate the effects caused by the different processing methods and may be established as a standard to optimize processing in digital holographic applications.

E. the benefits of applying the super-resolution technique

In Chapter 4, the super-resolution technique was introduced to improve the visualisation of large objects in digital holography. But the reconstructed SR hologram is not identical to the original high-resolution hologram. Therefore Zhang (2006) thought this might limit its application in measurement where precision is a concern.

A ball bearing with the radius of 7.95mm is placed off-axis and 355mm away from the CCD sensor. This stand-off distance is too close for our camera to be able to resolve the whole area of the ball bearing according to Eq. (4.14) so that it is beyond the resolvable angle of our camera in an off-axis geometry. The illumination angles in the exposures are -32.00° and -31.96° , respectively. For the same scene, holograms are taken by both normal mode and also by 4-scan mode. Fig. 5.15 shows the comparison of the reconstruction results produced using these two modes.

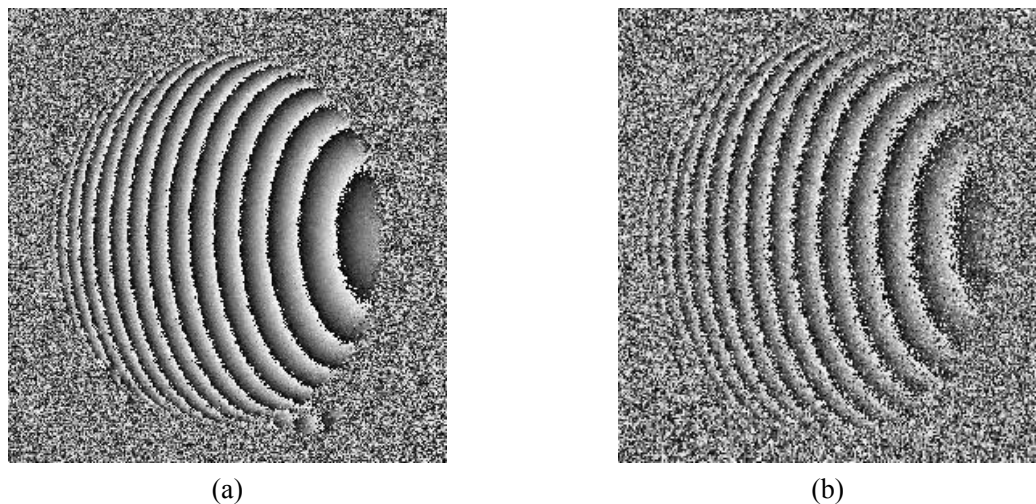


Fig. 5.15 The comparison of SR and non-SR holograms to measure the surface of a ball bearing by two-source contouring method: phase-difference images of (a) SR hologram and (b) non-SR hologram.

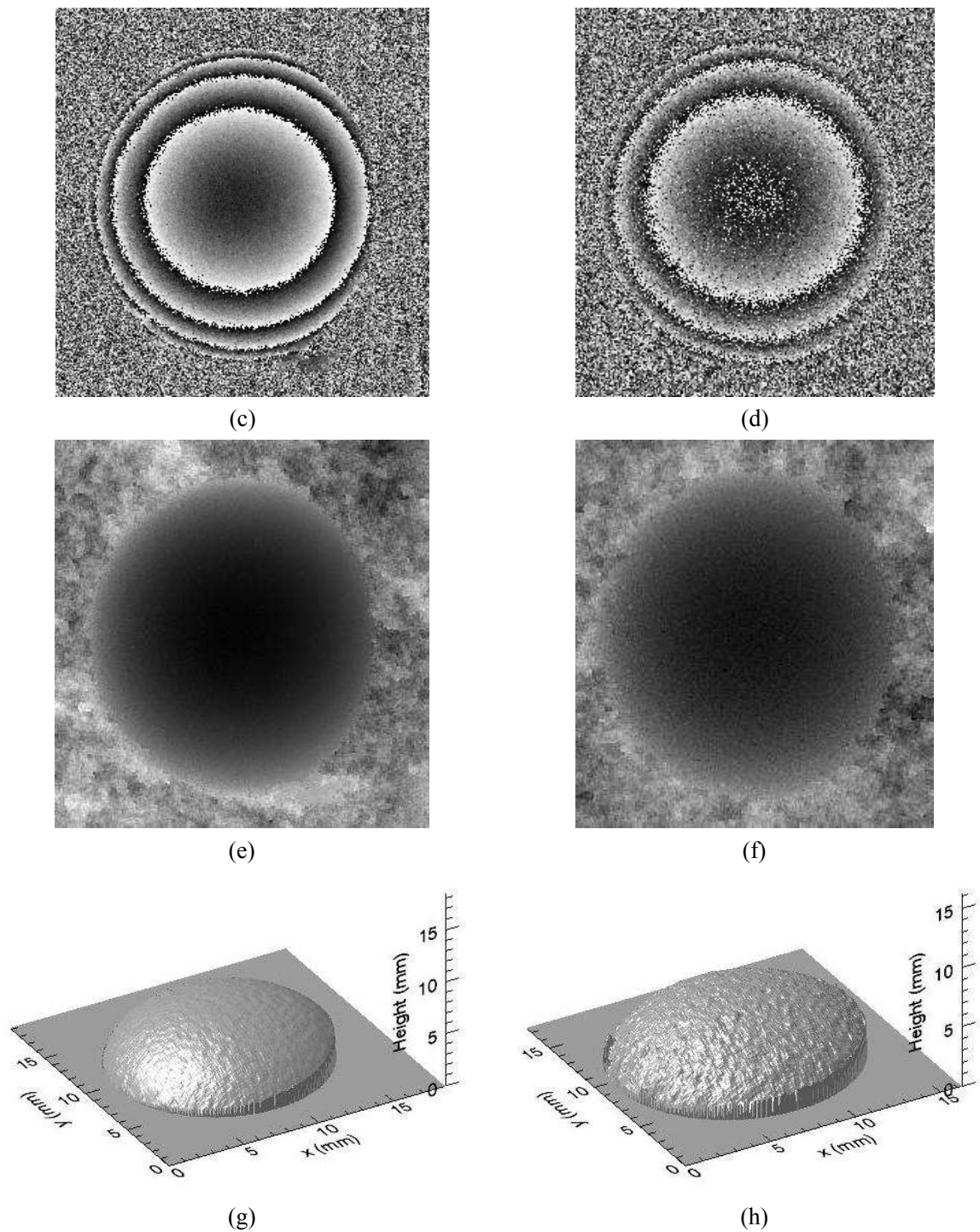


Fig. 5.15 (Continued) Phase-difference images removed from the linear factor of (c) SR hologram and (d) non-SR hologram; unwrapped phase-difference images of (e) SR hologram and (f) non SR holograms; reconstructed 3D surface of the roller (g) SR hologram and (h) non SR holograms.

The results shown in Fig. 5.15(a), (c), (e) and (g) were calculated from SR holograms with the resolution of 2048×2048 pixels, captured in 4-scan mode. Fig. 5.15(b), (d), (f) and (h) were calculated from LR holograms with the resolution of 1024×1024 pixels, captured in normal mode. It is clear in Fig. 5.15(b) that the surface contours of the ball bearing suffer from higher noise than that is shown in Fig. 5.15(a). In the reconstructed height

distribution image of Fig. 5.15(h), this can be seen most clearly. However for the images obtained using the super-resolution technique, i.e. captured via the 4-scan mode of the ProgRes® MF^{scan} camera, the phase data of the whole illuminated area of the roller is obtained. Furthermore, the 3D surface in Fig. 5.15(g) appears to have less noise than Fig. 5.15(h). Regarding the quantitative results, the measured radius of the ball bearing is $7.97 \pm 0.14 \text{ mm}$ determined by using a curve fitting technique, based on the Gradient-expansion algorithm, which represents an error of only 2% compared with the value of 7.95 mm as measured by a vernier calliper with an uncertainty of 0.02 mm . Therefore, the SR method is not only valid for visualisation application in digital holography, but are also valid for phase measurement, such as surface contouring as described here. This disagrees with Zhang's viewpoint (Zhang, 2006) and the accuracy of the measurement can be assured when applying SR methods in high-precision measurements of macroscopic objects.

F. the importance of the collimation of the reference wave

In digital holography, the reference wave in numerical reconstruction is a virtual wave designed to be equivalent to the one used in the recording procedure. In our experiments, a plane wave is used as the reference wave so that a plane wave is simulated for performing numerical reconstruction. Two beam expanders were used in our experiments. One is a compact $20\times$ beam expander made by Melles Griot, and the other is a combination of an objective ($NA = 0.62, f = 4 \text{ mm}$) and a collimation lens ($f = 75 \text{ mm}$). The design of mounting, and adjustment of the combination of the objective and the collimation lens, makes it difficult to get the desired optimized expanded collimated wave as the reference wave. This drawback was not revealed in the visualization application shown in the previous chapters. But in the phase measurements, such as two-source contouring method, it caused some problems such as those shown in Fig. 5.16.

It is noticeable that the contrast of the contours with a poorly collimated reference wave is much worse than that with a well collimated reference wave, while the speckle noise in Fig. 5.16(a) and (c) is more severe than it in Fig. 5.16(b) and (d). The higher order contours are submerged by the significant noise when the contour interval becomes small by increasing the change of the illumination angles $\Delta\theta$. In Fig. 5.16(c), the contours are hardly resolved

when $\Delta\theta = 0.03^\circ$. But for the same parameters with a well collimated reference wave, $\Delta\theta$ can be as high as 0.06° as shown in Fig. 5.14(a).

Therefore, careful attention must be paid to choosing and adjusting the beam expander in order to avoid undesired errors and achieve more reliable measurement results.

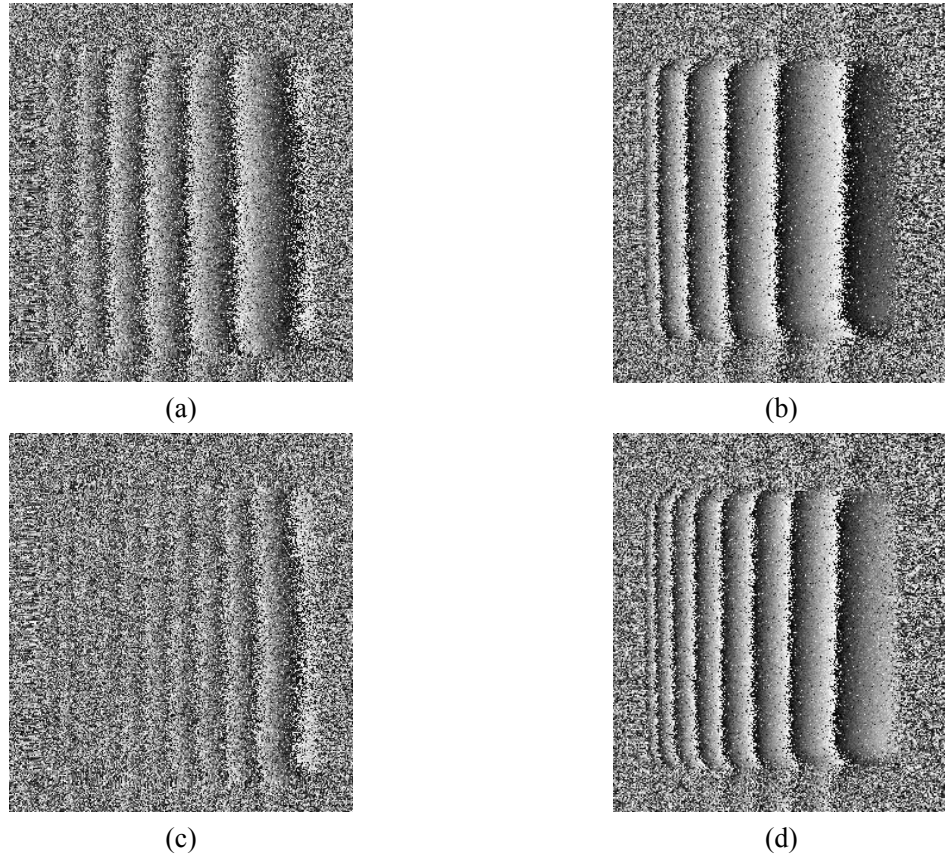


Fig. 5.16 The effect of the collimated reference wave to the directly subtracted phase-difference image in two-source method: (a) without and (b) with a well collimated reference wave when $\Delta\theta=0.02^\circ$; (c) without and (d) with a well collimated reference wave when $\Delta\theta=0.03^\circ$.

5.4 Two-wavelength methods

5.4.1 The conventional two-wavelength method

In this method, two holograms of the same object are recorded with two different wavelengths λ_1 and λ_2 . These holograms are recorded on the same holographic plate to form a composite hologram. When the composite hologram is illuminated by one of the recording wavelengths, say λ_1 , two images are reconstructed. The image recorded with wavelength λ_1 and reconstructed by λ_1 is an exact replica of the original object;

however, the other image that was recorded with wavelength λ_2 but reconstructed by λ_1 is shifted in its observation direction with respect to the original surface (Meier, 1965). These two reconstructed images interfere with each other to generate surface contours of the object. The general recording and reconstruction geometries are shown in Fig. 5.17.

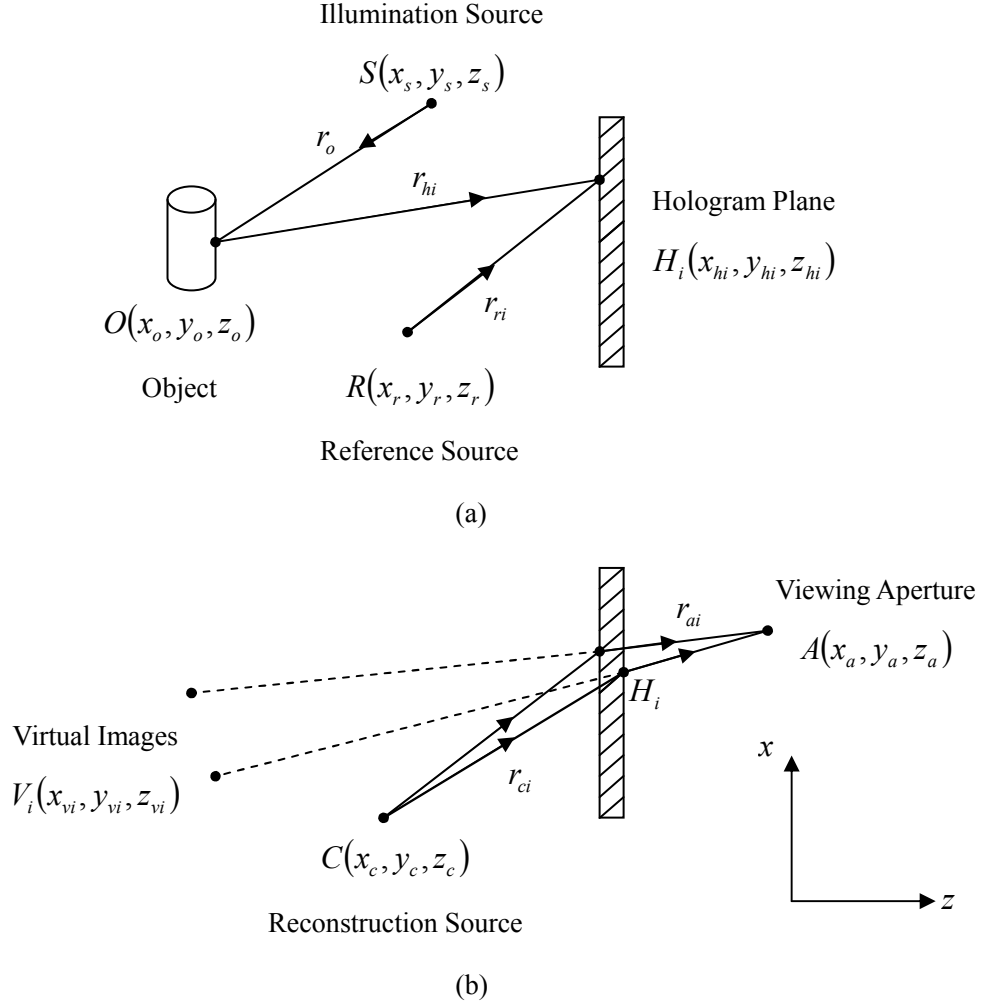


Fig. 5.17 General recording and reconstruction geometries: (a) recording; (b) reconstruction.

The equation for a constant phase difference is:

$$\Delta\varphi(x_o, y_o, z_o) = \varphi_1 - \varphi_2 = 2n\pi \quad (5.14)$$

which describes a surface in object space. The intersection of this surface with the object is the n th bright contour. Phase terms φ_1 and φ_2 represent the phase across the aperture due to the two points on the reconstructed virtual images according to the different recording wavelengths. Each of these consists of three components:

$$\varphi_i = \varphi_{ci} + \varphi_{hi} + \varphi_{vi} \quad (5.15)$$

where φ_{ci} is the phase of the reconstruction wave at the hologram, φ_{hi} is the phase contribution due to modulation by the hologram, and φ_{vi} is the phase contribution due to the path difference between the hologram and the aperture. These three components can be represented by the coordinate system in Fig. 5.17. When all the terms are introduced to Eq. (5.14) and the equation is simplified (Friesem and Levy, 1976), the basic relation for the analysis of two-wavelength contour holography is given by

$$\Delta\varphi = (r_o + r_h - r_r)\Delta k = 2n\pi \quad (5.16)$$

where $r_h = (r_{h1} + r_{h2})/2$, $r_r = (r_{r1} + r_{r2})/2$ and $\Delta k = k_1 - k_2 = 2\pi(\lambda_2 - \lambda_1)/(\lambda_1\lambda_2)$. Eq. (5.16) provides solutions for a family of surfaces in space.

For Fresnel hologram recording and readout arrangements, when the illumination, reference, and reconstruction beams are plane waves, if the illumination beam has an orientation of angle θ_o with respect to z axis, and the angular orientations of both reference beam and reconstruction beam are the same as θ with respect to z axis, Eq. (5.16) leads to

$$\Delta\varphi = [z_o(1 + \cos \theta_o) + x_o(\sin \theta_o - \beta \sin \theta)]\Delta k = 2n\pi \quad (5.17)$$

where $\beta = z_a/(z_a - \bar{z}_o)$, \bar{z}_o is the average distance between the hologram plane and the object plane. According to this equation, the family of planes are inclined at an angle θ_{inc} with respect to the hologram plane:

$$\tan \theta_{inc} = -\frac{\sin \theta_o - \beta \sin \theta}{1 + \cos \theta_o} \quad (5.18)$$

To eliminate the inclination, a recording and reconstruction geometry is chosen so that

$$\sin \theta_o - \beta \sin \theta = 0 \quad (5.19)$$

The depth resolution, i.e. the axial distance between adjacent fringes is thus given by

$$\Delta z_o = \frac{\lambda_1 \lambda_2}{(\lambda_2 - \lambda_1)(1 + \cos \theta_o)} \quad (5.20)$$

5.4.2 The two-wavelength method in digital holography

The recording procedure for digital holographic two-wavelength contouring is similar to

that of the conventional method, consisting of the recording of two holograms with different wavelengths λ_1 and λ_2 . Both holograms are recorded by the camera and saved in the computer. In the numerical reconstruction, these holograms are reconstructed separately using the true recording wavelengths, rather than only one of the recording wavelengths as is the case in the reconstruction procedure in the conventional method. This means that the hologram that had a recording wavelength λ_1 is reconstructed at wavelength λ_1 and the hologram that was produced using a recording wavelength λ_2 is reconstructed at wavelength λ_2 . Two resulting complex amplitudes $\Gamma_{\lambda_1}(\xi, \eta)$ and $\Gamma_{\lambda_2}(\xi, \eta)$ can therefore be correctly obtained. The phase is calculated as:

$$\varphi_{\lambda_1}(\xi, \eta) = \arctan \frac{\text{Im} \Gamma_{\lambda_1}(\xi, \eta)}{\text{Re} \Gamma_{\lambda_1}(\xi, \eta)}, \quad \varphi_{\lambda_2}(\xi, \eta) = \arctan \frac{\text{Im} \Gamma_{\lambda_2}(\xi, \eta)}{\text{Re} \Gamma_{\lambda_2}(\xi, \eta)} \quad (5.21)$$

As illustrated in section 5.3.2, both phase-contrast images are modulo 2π phase images. The phase difference should be calculated as in Eq. (5.7):

$$\Delta\varphi_{\lambda}(x, -z_0) = \begin{cases} \varphi_{\lambda_2}(x, -z_0) - \varphi_{\lambda_1}(x, -z_0), & \text{if } \varphi_{\lambda_2}(x, -z_0) \geq \varphi_{\lambda_1}(x, -z_0) \\ \varphi_{\lambda_2}(x, -z_0) - \varphi_{\lambda_1}(x, -z_0) + 2\pi, & \text{if } \varphi_{\lambda_2}(x, -z_0) < \varphi_{\lambda_1}(x, -z_0) \end{cases} \quad (5.22)$$

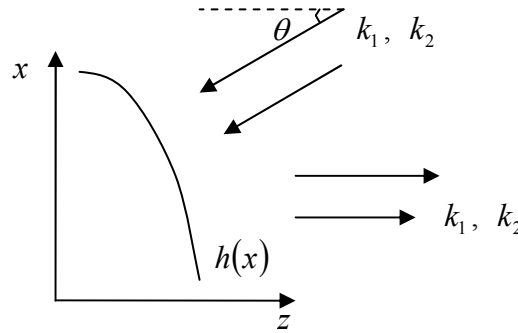


Fig. 5.18 Principles of surface contouring by two-wavelength digital holography

For surface contouring, if two holograms are recorded by tilt illumination as shown in Fig. 5.18, the phase terms φ_{λ_1} and φ_{λ_2} can be written as:

$$\varphi_{\lambda_1}(x, -z_0) = (k_1 \cos \theta + k_1)h(x) + k_1 x \sin \theta \quad (5.23)$$

$$\varphi_{\lambda_2}(x, -z_0) = (k_2 \cos \theta + k_2)h(x) + k_2 x \sin \theta \quad (5.24)$$

where $k_1 = 2\pi/\lambda_1$, $k_2 = 2\pi/\lambda_2$, and θ is the illumination angle with respect to z axis. Therefore the phase difference is

$$\Delta\varphi_\lambda(x, -z_0) = (k_2 - k_1)(1 + \cos\theta)h(x) + (k_2 - k_1)x \sin\theta \quad (5.25)$$

where it is assumed that the incident plane is parallel to the xz plane and the reference plane is the xy plane. The first term on the right-hand side of Eq. (5.25) includes the information of the surface height distribution, and the second term is a linear carrier introduced by the tilt illumination, which is similar to the two-source method in Eq. (5.5). This linear carrier can be eliminated either by numerical subtraction or by employing normal incidence ($\theta = 0$) in the recording procedure. For this case the phase difference is given by

$$\Delta\varphi_\lambda(x, -z_0) = (k_2 - k_1)(1 + \cos\theta)h(x) \quad (5.26)$$

The contour interval, i.e. the height difference corresponding to a phase difference of 2π , is:

$$\Delta h = \frac{\lambda_2 \lambda_1}{|\lambda_1 - \lambda_2|(1 + \cos\theta)} \quad (5.27)$$

For normal incidence,

$$\Delta h = \frac{\lambda_2 \lambda_1}{2|\lambda_1 - \lambda_2|} = \frac{\Lambda}{2} \quad (5.28)$$

where $\Lambda = \lambda_2 \lambda_1 / |\lambda_1 - \lambda_2|$ is the synthetic wavelength of λ_1 and λ_2 .

Although Eq. (5.20) for the conventional method is the same as Eq. (5.27), they are achieved by a different process. For the conventional method, it is hard to accurately determine the relationship between the observed contouring pattern and the surface profile of the object. This is because the image reconstructed by a wavelength different from the recording wavelength varies not only in phase, but also in position and magnification (Meier, 1965). When the difference of the two recording wavelengths is relatively large, the difference in image position will result in poor visibility of the contour fringes. In order to obtain accurate surface contours for the case of large wavelength differences, either there must be a change in the geometry of the holographic recording arrangement for each wavelength, or auxiliary optical components may be inserted during the recording procedure (Varner, 1971). This also causes some difficulty in performing the mathematical analysis precisely. The result presented in Eq. (5.20) is obtained as a result of certain simplifications (Friesem and Levy, 1976). However, all of these problems do not occur in

digital holography, as both holograms are reconstructed by the original recording wavelengths. No change in position and magnification is introduced by the reconstruction process. Eq. (5.27) is obtained strictly from the mathematical deduction. The result is in accordance with practice. Meanwhile, when two wavelengths are used that have a large difference between them, achromatic aberration must be considered when constructing the experimental setup.

5.4.3 Experimental results when using the two-wavelength method

An experimental system similar to that shown in Fig. 5.3 was constructed to investigate the two-wavelength contouring method. A 25mw tunable diode laser with a central wavelength of 685nm (TEC-100-0685-25 by Sacher Lasertechnik Group) was used as the light source. This was controlled by an MLD 1000 controller to vary the wavelength in the hologram recording process. The wavelength of the tunable diode laser was monitored by a Bristol LM621 wavelength meter provided by Bristol Instruments, Inc. A 685nm fibre optical isolator (FOI-01-13-685-4/125-S-40-3A-1-1-35 provided by OZ Optics Ltd.) was used to prevent back reflections into the diode laser system, which can cause wavelength instability of the diode laser. The laser beam coming out of this isolator was split into two beams by a 10/90 2×2 fibre coupler (FC632-90B-3FC-1APC manufactured by Thorlabs, Inc.), the weaker one of which was used as the reference wave and the stronger one of which was used to illuminate the object. In this application, the ProgRes[®] MF^{scan} camera was also used to record the digital holograms, as shown in Fig. 5.3. The measurement object was the roller with both a diameter and height of 16mm.

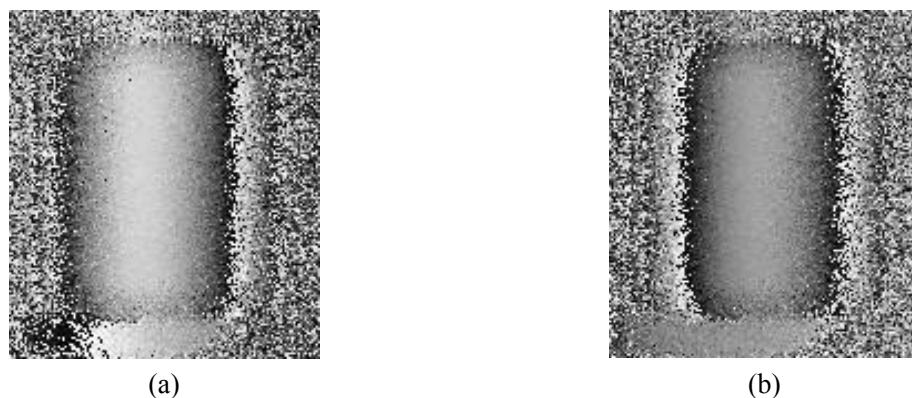
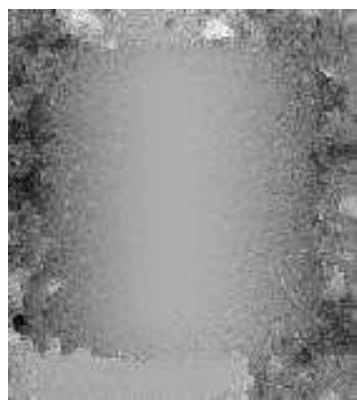
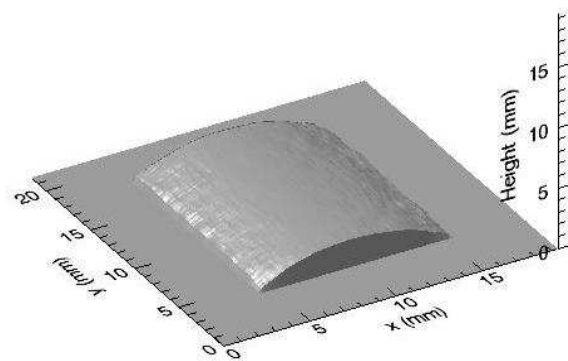


Fig. 5.19 Experimental results for the two-wavelength contouring method with a neighbouring contour interval of 2.41mm: (a) directly subtracted phase-difference image for the roller; (b) phase-difference image with the inclination factor removed.



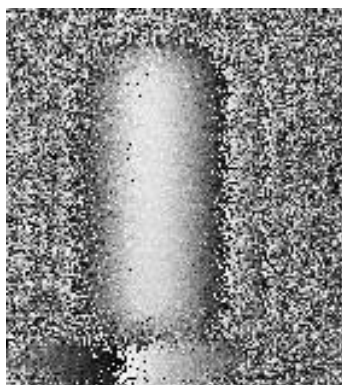
(c)



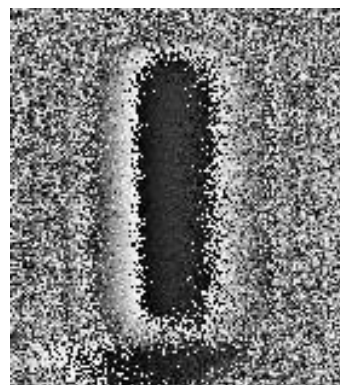
(d)

Fig. 5.19 (Continued) (c) unwrapped phase-difference image of (b); (d) 3D height distribution of the roller smoothed by a 7×7 window.

Fig. 5.19 shows an example of the two-wavelength contouring method. The roller was placed at a distance of 672mm from the CCD sensor. The roller was illuminated at an angle of -18° from the side. The wavelengths used in both exposures were 685.80nm and 685.70nm , respectively. Therefore the contour interval here according to Eq. (5.27) is 2.41mm . By using curve fitting, the measured radius of cylindrical roller in Fig. 5.19(d) is 8.70mm , 8% larger than the independently determined, via vernier caliper measurement, value of 8mm . Because of the shadowing problem in oblique illumination and the submersion of the finer contours in the speckle noise, only part of the surface of the roller is detected. This can be verified by changing the neighbouring contour interval shown in Fig. 5.20. With a decrease in the contour interval, i.e. an increase of the number of the contours, the detected area of the object shrinks because the finer contours hide in the noise of the phase-difference image. In Fig. 5.20(d), the measured radius of the cylindrical roller is 8.73mm .

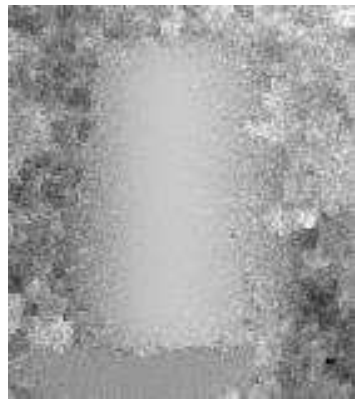


(a)

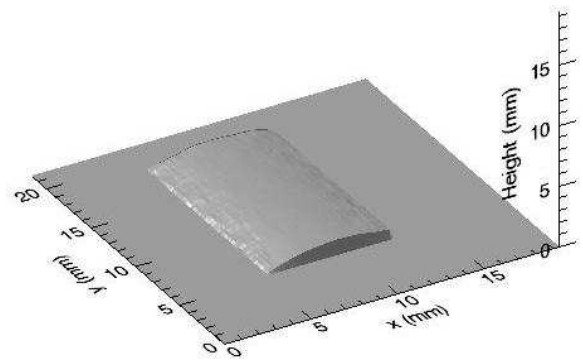


(b)

Fig. 5.20 Results of changing the contour interval to 1.34mm : (a) directly subtracted phase-difference image of the roller; (b) phase-difference image with the inclination factor removed.



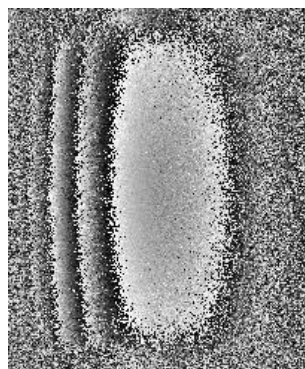
(c)



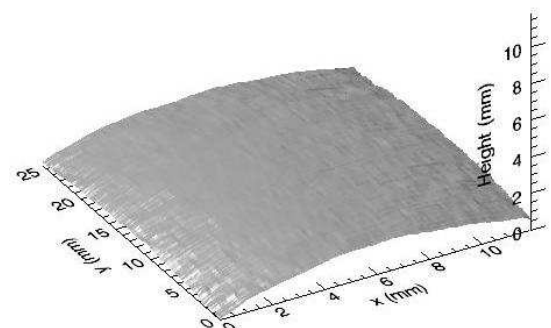
(d)

Fig. 5.20 (Continued) (c) unwrapped phase-difference image of (b); (d) 3D height distribution of the roller smoothed by a 7×7 window.

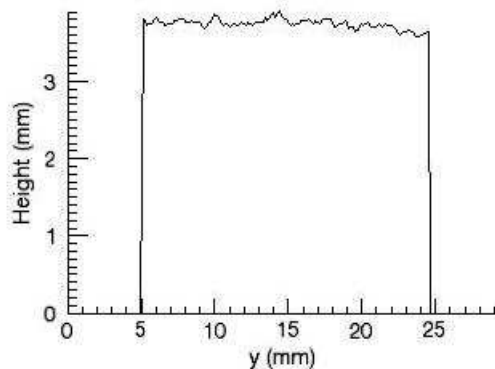
To ensure an accurate contouring result, a plane wave should be used for illumination so that the contour surfaces are almost plane and normal to the sensitivity vector. In our experimental geometry, the plane wave is achieved by adding a collimating lens after the output end of the fibre coupler. If this collimating lens is removed, a spherical wavefront is used to illuminate the object and this would, in general, cause distortion.



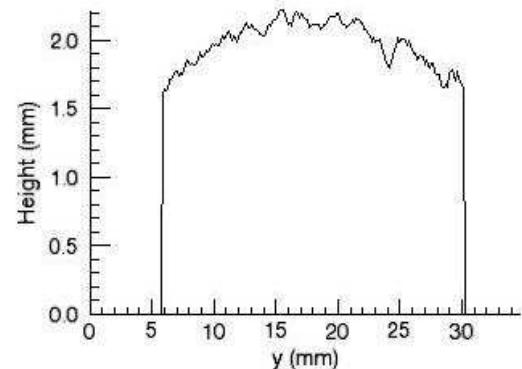
(a)



(b)



(c)



(d)

Fig. 5.21 (a) phase-difference image by illumination with a spherical wavefront; (b) reconstructed 3D height distribution of the roller; (c) the 1D height plot of the central line along the y axis in Fig. 5.19(d); (d) the 1D height plot of the central line along the y axis in Fig. 5.21(b).

In Fig. 5.21(a), the illumination by a spherical wavefront results in a closed contour which is an incorrect result for the object as it is a roller. Fig. 5.21(c) and (d) display the 1D height plot of the central line along the y axis of Fig. 5.19(d) and Fig. 5.21(b), respectively. Comparing them, it is obvious that the result obtained from the illumination by a spherical wavefront is distorted. In conventional holographic interferometry, though a divergent source is used to illuminate the object, it is placed quite far away from so that the distortion is small enough to be ignored. However, in our experimental geometry, the point source cannot be positioned far enough away from the object because of the limited output power available from the optical fibre. Therefore, according to our experience, it is recommended to avoid illuminating object with a spherical wavefront in optical contouring.

5.5 Two-refractive-index methods

5.5.1 The conventional two-refractive-index method

In this method, the object is mounted inside a cell with a plane-glass window (Varner, 1974). By changing the refractive index of the fluid around the object between the two exposures of the hologram, interference fringes are produced which contour the distance from the object to the window. The interpretation of the fringes is much simpler than in the previous technique, since a physical reference plane is provided by the cell window. The process is very similar to the use of holography for flow visualization. In visualization of two-dimensional flow fields the fringes contour a variable change in refractive index over a constant path length (Zelenka and Varner, 1969). In contouring, a variable path length is contoured for a constant change in refractive index as:

$$\Delta\phi(x, y) = \frac{2\pi}{\lambda}(n_1 - n_2)h \quad (5.29)$$

where n_1 and n_2 are the two refractive indices of the fluid surrounding the object during each of the two exposures, $\Delta\phi(x, y)$ is the phase difference, and h is the surface height distribution. The fringe orders are given by

$$N = \frac{d}{\lambda}(n_1 - n_2) \quad (5.30)$$

The contour interval is simply the change in surface height Δh which corresponds to an increment of one in the fringe order, that is

$$\Delta h = \frac{\lambda}{n_1 - n_2} \quad (5.31)$$

By suitable choice of the fluids, it is possible to obtain a wide range of contour intervals. With various liquids, contour intervals between $50\mu\text{m}$ and $500\mu\text{m}$ can be produced.

5.5.2 The two-refractive-index method in digital holography

Fig. 5.22 is a typical arrangement for contouring diffuse objects by the two-refractive-index method in digital holography. The plane of the window of the cell is kept normal to the mean direction of propagation of the object wave reaching the face plate of the sensor, so that the angle of incidence of the object beam and reference beam does not change. This is essential to avoid a lateral displacement or change of scale of the reconstructed image. In this experimental condition, the fringes are localized on the surface of the object. This condition also minimizes aberrations caused by imperfections in the material of the cell.

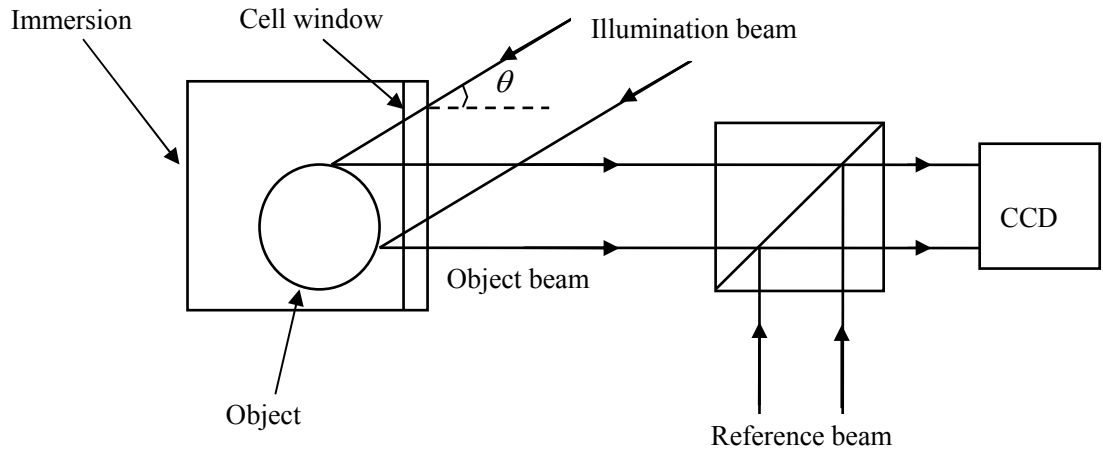


Fig. 5.22 The typical arrangement for contouring diffuse objects by the two-refractive-index method in digital holography

In this method, the object is located in an immersion cell containing a specific solution. By changing the concentration of the solution, the refractive index of the liquid is changed accordingly. Two holograms are captured at two different refractive indexes and these are numerically reconstructed to generate a phase-difference image $\Delta\phi$. The reconstructed phase images of refractive index n_1 and n_2 are given by:

$$\phi_{n1}(x, -z_0) = n_1 k \cos \theta_1 h(x) + n_1 k h(x) + n_1 k x \sin \theta_1 \quad (5.32)$$

$$\varphi_{n_2}(x, -z_0) = n_2 k \cos \theta_2 h(x) + n_2 k h(x) + n_2 k x \sin \theta_2 \quad (5.33)$$

where $\theta_1 = \arcsin(\sin \theta / n_1)$ and $\theta_2 = \arcsin(\sin \theta / n_2)$, θ is the incidence angle of the illumination beam as shown in Fig. 5.22. The phase difference $\Delta \varphi_n$ is given by subtracting φ_{n_1} from φ_{n_2} :

$$\Delta \varphi(x, -z_0) = k(n_2 - n_1)h(x) + k(n_2 \cos \theta_2 - n_1 \cos \theta_1)h(x) + kx(n_2 \sin \theta_2 - n_1 \sin \theta_1) \quad (5.34)$$

Then the contour interval can be deduced from Eq. (5.36) as given by:

$$\Delta h = \frac{\lambda}{(1 + \cos \theta_2)n_2 - (1 + \cos \theta_1)n_1} \quad (5.35)$$

In the experiment, two recorded digital holograms are reconstructed by the numerical methods described in Chapter 2 to generate two phase-contrast images. The related phase-difference image is calculated by

$$\Delta \varphi_n(x, -z_0) = \begin{cases} \varphi_{n_2}(x, -z_0) - \varphi_{n_1}(x, -z_0), & \text{if } \varphi_{n_2}(x, -z_0) \geq \varphi_{n_1}(x, -z_0) \\ \varphi_{n_2}(x, -z_0) - \varphi_{n_1}(x, -z_0) + 2\pi, & \text{if } \varphi_{n_2}(x, -z_0) < \varphi_{n_1}(x, -z_0) \end{cases} \quad (5.36)$$

which is similar to Eq. (5.7). By applying a phase unwrapping algorithm to this phase-difference image, the phase map of the detected surface is produced and can be converted to a surface height distribution.

Fig. 5.23 gives an example of the two-refractive-index technique that was provided by Hossain et al. (2007). An immersion cell of dimensions of $60\text{mm} \times 30\text{mm} \times 65\text{mm}$ having optically flat windows, with flatness of the order of $\sim \lambda/6$, was used in the experiment. The measured object was a nearly spherical steel ball of 20mm diameter immersed in the immersion cell. The refractive index of the water-ethanol solution was varied by adjusting the concentration of the ethanol in the water. The refractive indices were measured by an Abbe-Refractometer.

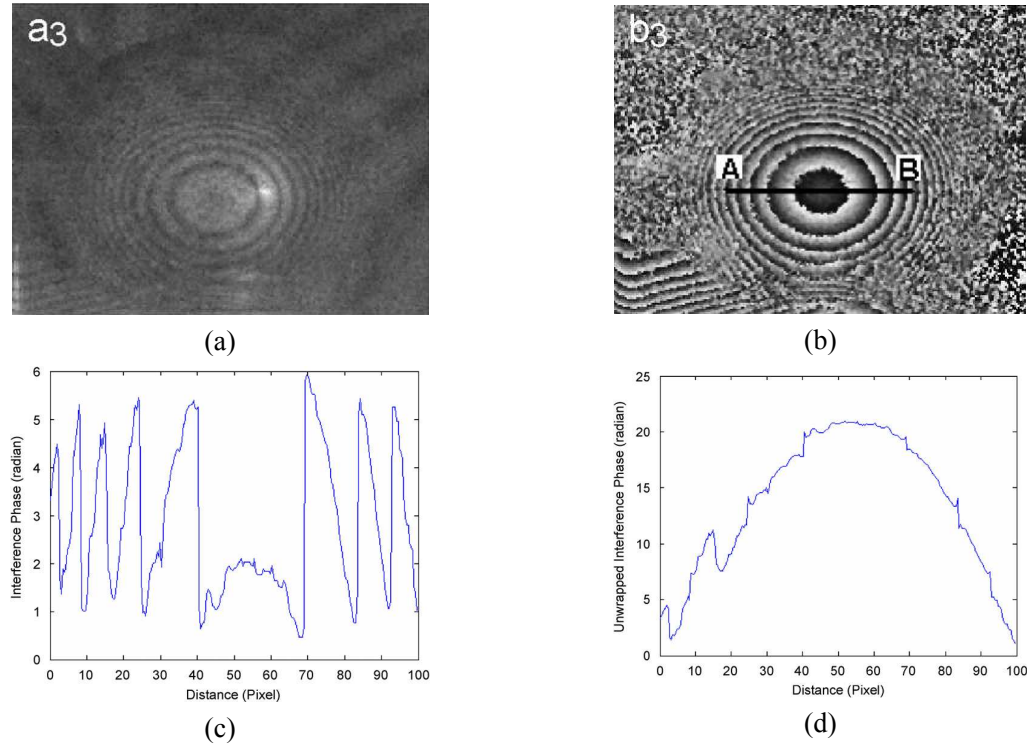


Fig. 5.23 An example of the two-refractive-index technique: (a) the phase-difference image reconstructed from two Fourier holograms corresponding to the distilled water – ethanol solutions of different concentrations (different refractive indices) (1.3322, 1.3332); (b) the phase-difference image after removal of the zero-order term and 4×4 median filtering; (c) one-dimensional wrapped phase along the length of the line AB in (b); (d) unwrapped phase of (c) (Reprinted from (Hossain et al., 2007). Copyright (2007), with permission from Elsevier).

The depth contour interval in this example is 0.29mm . It decreases with an increase of the difference in refractive indices, according to Eq. (5.37). Quantitative information regarding the depth of the spherical ball can be extracted via the unwrapped phase map shown in Fig. 5.23(d) and the knowledge of the depth contour interval.

5.6 Decorrelation effects in optical contouring

In the derivations of equations about optical contouring, it has been assumed that the random amplitude and phase of an individual speckle remains unchanged in the course of capturing two holograms which to form the fringes. The amplitudes in the recording plane in both recording states contain terms of the form

$$U_1 = u_s \exp i\varphi_s \quad (5.37)$$

$$U_2 = u'_s \exp[i(\varphi'_s + \Delta\varphi)] \quad (5.38)$$

where u_s and u'_s are the real amplitude of the object wave, φ_s and φ'_s are the phase of the object wave in both recording states, $\Delta\varphi$ is the change of the phase between both exposures. And it has been assumed that $u_s = u'_s$ and $\varphi_s = \varphi'_s$, so that the amplitudes are fully correlated when $\Delta\varphi = 2n\pi$. However, if $u_s \neq u'_s$ and $\varphi_s \neq \varphi'_s$, the maximum value of the correlation coefficient of U_1 and U_2 is reduced. If the change is sufficiently large, the speckle patterns will be decorrelated and no fringes will be observed. Decorrelation effects are related to surface roughness, the NA and the tilt angle. Detailed discussion about the decoorelation effects can be found in the literature (Jones and Wykes, 1983).

Our experiments are carried out on an optical bench. A Michelson interferometer was constructed to test the stability of the environment. Noticeable movement of the fringes on the paper screen can be observed. It mainly caused by the heat convection and the ambient vibration in the building. To reduce the heat convection, the experimental system is covered by a large box-like enclosure. Meanwhile, an air-conditioning unit was installed to control the room temperature. To prevent the ambient vibration in the building, the optical bench should be equipped with vibration isolation function. But limited by our experimental condition, this cannot be resolved except by buying a new vibration isolation bench. Therefore, it can be discerned that the full potential of two-source and two-wavelength contouring techniques might have not been fully exploited. However from the obtained results, the decorrelation effect would not appear to be significant.

5.7 Conclusions

In this chapter, three optical contouring methods have been discussed in detail. These are the two-source contouring method, the two-wavelength contouring method, and the two-refractive-index contouring method. Limited by experimental conditions, the two-refractive-index technique has not been demonstrated, but the other two techniques have been verified by our experiments. Through the results presented in this chapter, some conclusions can be drawn as follows.

The two-source technique is theoretically the simplest contouring method. The idea is quite straightforward and the experimental system is not complicated. The contour interval is strongly dependent on the change of both illumination angles, $\Delta\theta$. The shadowing problem is unavoidable in the two-source method because of the oblique illumination angle of the

object. Careful arrangement of the positions of both the object and the reference wave is required in order to prevent the object obstructing the reference beam. An increase in $\Delta\theta$ results in a decrease in the contour interval. The choice of contour interval affects the quality and accuracy of the final results. Smaller contour intervals are beneficial for measuring steeper surfaces, but the limited resolution of the CCD device and decorrelation effects, related to the tilt angle, also lead to poor contrast in the contouring map. Special care must be taken to form a collimated plane reference wave in the recording procedure; otherwise the contour map is of poor quality and fails to extract the correct surface information.

For the two-wavelength technique, no shadowing problem exists for proper arrangements and there is no requirement for any moving parts during the two exposures. However a laser source providing more than one wavelength is needed to apply this technique. The stability of both the wavelength and the intensity during the exposure time must be good enough to capture holograms with sufficient quality to allow reconstruction of the height distribution of the surface. The contour interval of this technique is controlled by the recording wavelengths used in both exposures. Larger differences between these two wavelengths provide smaller height sensitivities and the ability to measure surfaces with finer structures. But it should be noted that too small a contour interval might result in too many contours which cannot be resolved by the CCD camera. The tunable laser system used in our experiments proved to have quite poor wavelength stability. Additionally, though the use of optical fiber components to deliver the laser beam ensures good spatial filtering for imaging, this configuration means that the laser beam used to illuminate the surface under study is attenuated to only 10% of the original output power. Hence both the noise from the camera and the laser speckle noise significantly degrade the quality of the contouring image, as shown in Fig. 5.19 and Fig. 5.20.

In the two-refractive-index techniques, the object is mounted inside a cell with a plane-glass window. The object is immersed in fluid. The whole equipment is harder to handle practically because of this use of liquid. Moreover, the measurement of the refractive index of the liquid is also complicated. The contour interval of this technique is determined by the refractive index of the liquid and its subsequent change.

Plane-wave reference and object beams are used for optical contouring techniques, because they allow for a uniform treatment of all the object points. In the two-source contouring

technique, another advantage of using these plane-wave beams is the fact that the necessary adjustment can be simplified in the form of angular rotation of the reference and object beams.

This is the first time that the super-resolution method has been tested in its application to contouring techniques and it provides measurement results for objects which are larger than those which can be resolved with the CCD camera. The relevant results are given in Fig. 5.15. The holograms obtained from super-resolution method provide a less noisy 3D surface profile with similar accuracy to holograms obtained without using super-resolution methods. Considering the comparisons of the quantitative results shown in sections 5.3.4 C and D, this experimental two-source contouring approach may be used as a benchmark to evaluate the effects of the different processing techniques in digital holography.

References:

- BEVINGTON, P. R. & ROBINSON, D. K. (1992) *Data reduction and error analysis for the physical sciences*, McGraw-Hill, Inc.
- CUCHE, E., BOVILACQUA, F. & DEPEURSIAGE, C. (1999) Digital holography for quantitative phase-contrast imaging. *Optics Letters*, 24, 291-3.
- FRIESEM, A. A. & LEVY, U. (1976) Fringe formation in two-wavelength contour holography. *Applied Optics*, 15, 3009-20.
- GHIGLIA, D. C. & PRITT, M. D. (1998) *Two-dimensional phase unwrapping: theory, algorithms, and software*, John Wiley & Sons, Inc.
- HILDEBRAND, B. P. & HAINES, K. A. (1967) Multiple-wavelength and multiple-source holography applied to contour generation. *Journal Of The Optical Society Of America*, 57, 155 - 162.
- HOSSAIN, M. M., SHEORAN, G., MEHTA, D. S. & SHAKHER, C. (2007) Contouring of diffused objects by using digital holography. *Optics and Lasers in Engineering*, 45, 684-689.
- JONES, R. & WYKES, C. (1983) *Holographic and Speckle Interferometry*, Cambridge University Press.
- MEIER, R. W. (1965) Magnification and third-order aberrations in holography. *Journal Of The Optical Society Of America*, 55, 987 - 992.
- PALACIOS, F., GONCALVES, E., RICARDO, J. & VALIN, J. L. (2004) Adaptive filter to improve the performance of phase-unwrapping in digital holography. *Optics Communications*, 238, 245-51.
- SCHNARS, U. (1994) Direct phase determination in hologram interferometry with use of digitally recorded holograms. *Journal of the Optical Society of America A (Optics, Image Science and Vision)*, 11, 2011-15.
- VARNER, J. R. (1971) Simplified multiple-frequency holographic contouring. *Applied Optics*, 10, 212 - 213.
- VARNER, J. R. (1974) Holographic and Moire surface contouring. IN ERF, R. K. (Ed.) *Holographic non-destructive testing*. Academic Press.
- YAMAGUCHI, I., OHTA, S. & KATO, J. (2001) Surface contouring by phase-shifting

digital holography. *Optics and Lasers in Engineering*, 36, 417-28.

ZELENKA, J. S. & VARNER, J. R. (1969) Multiple-index holography contouring. *Applied Optics*, 8, 1431-4.

ZHANG, S. (2006) Application of super-resolution image reconstruction to digital holography. *Eurasip Journal on Applied Signal Processing*, 2006, 1-7.

6 Conclusions and the Future Work

6.1 Conclusions

For the measurement of macroscopic objects over long distances, say tens of millimetres, Fresnel approximation offers the best option. But for microscopic objects, the convolution approach is more competitive than the Fresnel approximation algorithm, especially when the distance between the object and the CCD sensor is too small to satisfy the requirements of Fresnel diffraction.

In digital holography, the zero-order term and the twin image degrade the quality of the reconstruction results more seriously than they do in conventional holography. It is essential to suppress them in order to produce correct image information. Spatial filtering is very effective for removing the unwanted terms for off-axis holograms.

Despite the increased system complexity, phase-shifting digital holography provides the best effect for suppressing the influence of the unwanted terms for in-line holography. By taking multiple holograms with different phase shifts in the reference wave, the resolution of the CCD device can be fully exploited.

As a technique that is based on interferometry, the reference wave and the object wave should be kept at comparable intensities in order to achieve good quality reconstruction results.

Speckle noise is unavoidable and very troublesome in digital holography when dealing with objects with rough surfaces. Digital filters can be applied to the holograms in order to reduce the speckle noise. However, the use of multiple holograms to reduce the speckle noise provides more satisfactory results and is more commonly employed. The multiple holograms are obtained by varying the speckle pattern of the object wave. This can be achieved by introducing some movement between different exposures in order to produce holograms with various speckle patterns.

The super-resolution image reconstruction approach has been incorporated into our measurement system by the use of a camera, the Jenoptik ProgRes[®]MF^{scan}, with microscanning function. The introduction of this approach not only removes aliasing effects in order to make the measurements more accurate, but also enables improved quality of the reconstruction results and either much larger observation ranges or closer measurement distances for the same object.

Two-source, two-wavelength and two-refractive-index contouring techniques are proven approaches that have been used in digital holography. The measured height distribution correctly reflects the true shape of the surface under study, but noise levels are significant due to the limited size of the CCD sensor and the weak reflected light returning from the object's surface. Therefore, optical contouring of macroscopic objects can now be performed using digital holography, but it is not currently a good choice to perform macroscopic 3D surface measurements.

Plane-wave reference and object beams are used in optical contouring techniques because they allow for a uniform treatment of all object points. A spherical wave, used to illuminate the object's surface, leads to unwanted distortion of the measured 3D surface.

The super-resolution method has been tested in surface contouring techniques and it provides measurement results for surfaces which are larger than those which can be resolved by the CCD camera. The precision of its measurement is similar to the measurements without using super-resolution methods.

The quantitative results from two-source contouring of an object with accurately known dimensions can be used as a benchmark to evaluate the effects caused by the different processing methods and establishes a standard to optimize the processing in digital holographic applications.

6.2 Future work

Three suggestions for future work can be outlined for digital holography as listed below.

A. Quantitative benchmarking of different processing techniques and optimizations

As illustrated in 5.3.4, two-source contouring or other quantitative measurements can be used as a numerical benchmark to evaluate the different processing techniques. This work is implied in the comparisons in 5.3.4 (Table 5.1 ~ 5.4). However, further work needs to be done in the future to validate this idea.

B. Influence of the bit depth of the CCD camera in digital holography

By applying super-resolution methods to digital holography, higher spatial resolutions are achievable, which may either be used to measure the objects with increased size or alternatively to decrease the stand-off distance to the object. However from a consideration of the reconstruction results for the super-resolution method which are presented in Chapters 4 and 5, it can be seen that the restoration of the detail of the object has not been improved by the super-resolution methods.

According to the work of Skydan et al. (2003), an increase in the image bit depth may well be significantly beneficial in terms of increasing the accuracy of phase-measuring fringe analysis systems. Digital holographic contouring systems bear a resemblance to fringe analysis systems because they also produce a result in terms of a phase measurement. This leads us to consider that it might also be beneficial to the accuracy of digital holographic measurements by employing a CCD camera with an increased bit depth, i.e. one that is greater than the 14 bits provided by the ProgRes[@] MF^{scan} camera that was used in this research. Further work could therefore be performed in order to investigate the influence of the bit depth of the CCD camera in digital holography.

C. Digital holographic microscopy (DHM)

Although digital holography can be applied to the interferometric measurements of macroscopic objects, as illustrated in this thesis, especially via the introduction of either phase-shifting digital holography or by super-resolution techniques, which enable it to employ the full potential of the CCD, it is still confined by the intrinsic characteristics of the CCD sensor. Comparing digital holography to conventional holographic recording media, the CCD has much lower resolution which is still insufficient for the production of

high quality holograms for the macroscopic objects with a size of a few centimeters and at shorter measurement distances – typically less than $0.5m$. Meanwhile, the small size of the CCD chip also results in significant speckle noise, which has been shown in this thesis.

However in the microscopic field, digital holography has been a topic of intense recent interest that is being investigated by increasing numbers of researchers. The resolution of current CCD cameras is no longer a constraint because of the dramatically decreased size of the microscopic samples. Also there is no need to use a very powerful laser beam to illuminate the sample, as most microscopic samples are transparent. The greatest potential of digital holography when applied to microscopy, is its ability to perform numerical focusing. In microscopy, the depth of field of the system is very small due to the high magnifications involved. Therefore, in order to obtain 3D information about an object at microscopic resolutions, several refocusing steps are usually needed, e.g. the scanning process in confocal microscopic systems. However, all this 3D information can be recorded in a single hologram and certain layers of the object can be refocused by use of the numerical methods that were discussed in Chapter 2. Moreover, due to the fact that the size of the microscopic sample is comparable to the wavelength of the light source, the phase-contrast image that is obtained by the numerical reconstruction can be used directly to render the surface or other properties of the sample.

The lateral resolution of the reconstructed image produced by the Fresnel approximation algorithm in digital holography is described in Eq. (2.25). A small reconstruction distance is needed in order to achieve the desired lateral resolution. But this often violates the requirements for a Fresnel approximation. The distance does not affect the application of the convolution approach. According to Eq. (2.39), the lateral resolution of the convolution approach is equal to the pixel size of the CCD sensor. Unfortunately, the pixel size of state of the art cameras is about $5\mu m$ and so this is insufficient for microscopy.

To enhance the lateral resolution of digital holography, some modifications have been introduced. Rather than a plane wave, a spherical wave is used as the reference wave. Fig. 6.1 shows set-ups for DHM with and without optical fiber components. By use of optical fiber components, the whole geometry is more compact, easier to build and adjust. But the single mode fiber changes the polarization status of the delivered laser beam which causes

problems with the phase-shifting technique. Meanwhile, the intensity loss caused by the fiber coupling is typically higher than 60~70%.

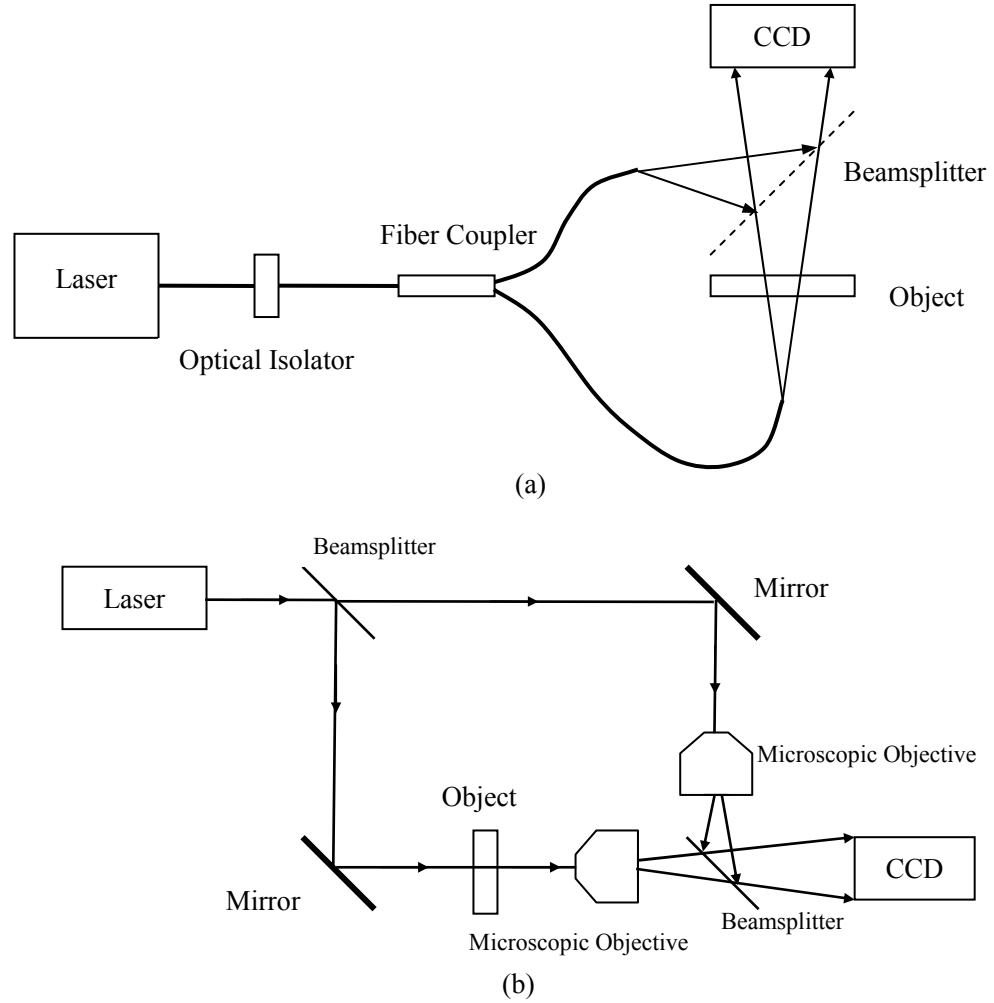


Fig. 6.1 Schematic diagram of a digital holographic microscope: (a) with optical fiber components; (b) without optical fiber components

For a hologram generated by a spherical reference wave, the lateral magnification of the reconstructed virtual image is (Schnars and Juptner, 2005):

$$M = \left[1 + \frac{d}{d'_r} \frac{\lambda_1}{\lambda_2} - \frac{d}{d_r} \right]^{-1} \quad (6.1)$$

where d_r and d'_r describe the respective distances between the source point of the spherical reference wave and the hologram plane in the recording and reconstruction process. λ_1 and λ_2 are the wavelengths used in the recording and reconstruction processes. The reconstruction distance d' , i.e. the distance between the reconstructed image and the hologram plane, can be calculated by (Meier, 1965):

$$d' = \left[\frac{1}{d_r'} + \frac{\lambda_2}{\lambda_1} \frac{1}{d} - \frac{1}{d_r} \frac{\lambda_2}{\lambda_1} \right]^{-1} \quad (6.2)$$

If the same reference wavefront is used for recording and reconstruction, it follows that $d' = d$. Magnification can be introduced by changing the wavelength or the position of the source point of the reference wave in the reconstruction process. In digital holography, the magnification can be easily introduced by changing the source point of the reference wave. If the desired magnification factor is determined, the reconstruction distance can be calculated by combination of Eqs. (6.1) and (6.2) with $\lambda_1 = \lambda_2$:

$$d' = d \cdot M \quad (6.3)$$

To enlarge the image the source point of the reference wave needs to be placed at a distance given by:

$$d_r' = \left[\frac{1}{d'} - \frac{1}{d} + \frac{1}{d_r} \right]^{-1} \quad (6.4)$$

The reference wave is now described by

$$R(x, y) = \exp \left(-i \frac{2\pi}{\lambda} \sqrt{d_r'^2 + (x - x_r')^2 + (y - y_r')^2} \right) \quad (6.5)$$

where $(x_r', y_r', -d_r')$ is the position of the source point of the reference wave in the reconstruction process.

The entire numerical reconstruction process of DHM may be summarized as follows: after determination of the desired magnification, the reconstruction distance d' is calculated using Eq. (6.3). Then the position of the source point d_r' of the spherical reference wave that is used for reconstruction is calculated by applying Eq. (6.4). The spherical reference wave can be calculated by using d_r' via Eq. (6.5). Finally one of the several numerical reconstruction algorithms that have been discussed in Chapter 2 may be applied in order to obtain the magnified amplitude and phase-contrast images for the object. Because of the aberrations introduced by the microscope objective in the reference wave, aberration correction measures must be performed in order to achieve reconstruction results that are free from such faults.

In summary, by applying digital holography to microscopy, this technology has great potential in diverse research areas such as metrology and quality control of micro-components and cell observations in biology.

References:

- MEIER, R. W. (1965) Magnification and third-order aberrations in holography. *Journal Of The Optical Society Of America*, 55, 987 - 992.
- SCHNARS, U. & JUPTNER, W. (2005) *Digital holography - digital hologram recording, numerical reconstruction, and related techniques*, Springer.
- SKYDAN, O. A., LILLEY, F., LALOR, M. J. & BURTON, D. R. (2003) Quantization error of CCD cameras and their influence on phase calculation in fringe pattern analysis. *Applied Optics*, 42, 5302-5307.

Bibliography

- BEVINGTON, P. R. & ROBINSON, D. K. (1992) *Data reduction and error analysis for the physical sciences*, McGraw-Hill, Inc.
- BRYNGDAHL, O. & WYROWSKI, F. (1990) Digital holography - computer generated holograms. *Progress in Optics*.
- CHEN, F., BROWN, G. M. & MUMIN, S. (2000) Overview of three-dimensional shape measurement using optical methods. *Optical Engineering*, 39, 10-22.
- GHIGLIA, D. C. & PRITT, M. D. (1998) *Two-dimensional phase unwrapping: theory, algorithms, and software*, John Wiley & Sons, Inc.
- HARIHARAN, P. (1984) *Optical holography*, Cambridge university press.
- LEE, W. H. (1978) *Computer-generated holograms: techniques and applications*.
- RASTOGI, P. K. (1997) *Optical measurement techniques and applications*, Boston, Artech House, INC.
- SCHNARS, U. & JUPTNER, W. (2005) *Digital holography - digital hologram recording, numerical reconstruction, and related techniques*, Springer.
- SMITH, H. M. (1969) *Principles of holography*, Wiley.

Appendix: some programs for our applications written in IDL

1. Program to generate the digital hologram of an object by calculating the diffraction pattern point by point

```

PRO FT_CGH
; This program is to generate a hologram by adding the diffraction fields of
every point source
; The unit of length is millimeter.
; The parameters are saved in a .dat file and the generated hologram is saved
in a .hol file
; The generated hologram is for FT reconstruction

; Input parameters of ccd
Xnum = 512 ; pixel number in x axis
Ynum = 512 ; pixel number in y axis
ccdsp = 0.01 ; the distance between two adjacent pixels in the CCD
ccdXlength = (Xnum - 1) * ccdsp
ccdYlength = (Ynum - 1) * ccdsp

; Input parameters of object field
Zdist = 500.0 ; the distance between CCD plane and the object plane
wavelength = 0.0006328
objXsp = wavelength * Zdist / (Xnum * ccdsp) ; the pixel size in the x axis
at the reconstructed object plane
objYsp = wavelength * Zdist / (Ynum * ccdsp) ; the pixel size in the y axis
at the reconstructed object plane
objXlength = (Xnum - 1) * objXsp
objYlength = (Ynum - 1) * objYsp

; Input parameters of reference beam
Ax = 0.
Ref_amp = 0.0

; Initiate some arrays
objPx = fltarr(Xnum)
objPy = fltarr(Ynum)
ccdPx = fltarr(Xnum)
ccdPy = fltarr(Ynum)
obj = fltarr(Xnum, Ynum)
obj_real = dblarr(Xnum, Ynum)
obj_imag = dblarr(Xnum, Ynum)
total_real = dblarr(Xnum, Ynum)
total_imag = dblarr(Xnum, Ynum)
line_real = fltarr(Xnum)
line_imag = fltarr(Xnum)
ref_real = fltarr(Xnum, Ynum)
ref_imag = fltarr(Xnum, Ynum)
ref_mean = dblarr(Xnum)

```

```

; Set up the orthogonal coordinates of object plane
objPx[0] = -1. * objXlength / 2.
for i=1, Xnum-1 do begin
    objPx[i] = objPx[i-1] + objXsp
endfor
objPy[0] = -1. * objYlength / 2.
for i=1, Ynum-1 do begin
    objPy[i] = objPy[i-1] + objYsp
endfor

; Set up the orthogonal coordinates of ccd plane
ccdPx[0] = -1. * ccdXlength / 2.
for i=1, Xnum-1 do begin
    ccdPx[i] = ccdPx[i-1] + ccdsp
endfor
ccdPy[0] = -1. * ccdYlength / 2.
for i=1, Ynum-1 do begin
    ccdPy[i] = ccdPy[i-1] + ccdsp
endfor

; Generate the object pattern at the object plane, which is a "F" letter in
this case
obj[106:166, 255:259] = 128
obj[106:156, 205:209] = 128
obj[106:110, 155:257] = 128
window, 0, xsize = 512, ysize = 512, title = 'Object'
tvsc1, congrid(obj, 512, 512)
obj = sqrt(obj) ; intensity of the object wave at the object plane
obj_phase = fltarr(Xnum, Ynum) ; phase of the object wave at the object plane
which is randomly distributed
seed = 11
err = randomn(seed, Xnum, Ynum)
obj_phase = err * !pi / 4

; Generate the object diffraction field
wavenum = 2. * !pi / wavelength
source_num = 01
t1 = systime(1)
Zeuclid = Zdist ^ 2.0
for u=0, Xnum-1 do begin ; for each point source in turn
    for v=0, Ynum-1 do begin
        if (obj[u,v] ne 0) then begin
            source_num = source_num + 1
            for x=0, Xnum-1 do begin ; for each CCD column
                Xeuclid = (ccdPx[x] - objPx[u]) ^ 2.0
                for y=0, Ynum-1 do begin ; for each CCD row
                    Yeuclid = (ccdPy[y] - objPy[v]) ^ 2.0 ; calculate the y
component of the euclidean distance
                    euclid = sqrt(Xeuclid + Yeuclid + Zeuclid)
                    ;arg = wavenum * euclid
                    arg = -1. * wavenum * euclid + 2 * obj_phase[u,v]
                    obj_real[x,y] = obj[u,v] * cos(arg) / euclid ; the real part
of the point source wavefront
                    obj_imag[x,y] = obj[u,v] * sin(arg) / euclid ; the imaginary
part of the point source wavefront
                endfor
            endfor
            total_real = total_real + obj_real
            total_imag = total_imag + obj_imag
            print, source_num, ' points have been completed.'
        endif
    endfor
endfor

```

```

        endfor
    endfor
    t2 = systime(1)
    print, 'It took ', t2-t1, ' seconds to calculate the diffraction field of the
    object.'
    temp_real = total_imag / wavelength
    temp_imag = total_real / wavelength
    total_real = temp_real
    total_imag = temp_imag
    obj_inten = total_real ^ 2.0 + total_imag ^ 2.0
    obj_phase = atan(total_imag, total_real)
    window, 1, xsize = 512, ysize = 512, title = 'Object Diffraction Intensity
    Field'
    tvscl, congrid(obj_inten, 512, 512)
    window, 2, xsize = 512, ysize = 512, title = 'Object Diffraction Phase Field'
    tvscl, congrid(obj_phase, 512, 512)

; Generate the reference light field
ref_amp = sqrt(mean(obj_inten))
for x=0, Xnum-1 do begin
    phase_angle = -1.0 * wavenum * sin(Ax) * ccdPx[x]
    line_real(x) = cos(phase_angle) * ref_amp
    line_imag(x) = sin(phase_angle) * ref_amp
endfor
for y=0, Ynum-1 do begin
    ref_real[,y] = line_real
    ref_imag[,y] = line_imag
endfor
ref_phase = atan(ref_imag, ref_real)
window, 3, xsize = 512, ysize = 512, title = 'Reference Beam Phase'
tvscl, congrid(ref_phase, 512, 512)

; Computer-generated Hologram
ccd_real = total_real + ref_real
ccd_imag = total_imag + ref_imag
ccd = float(ccd_real ^ 2.0 + ccd_imag ^ 2.0)
window, 4, xsize = 512, ysize = 512, title = 'Hologram'
tvscl, congrid(ccd, 512, 512)

; Save the hologram
file = dialog_pickfile(/write)
holofilename = file + '.hol'
openw, unit, holofilename, /get_lun
writeu, unit, ccd
close, unit

; Save the set up parameters
datafilename = file + '.dat'
openw, unit, datafilename, /get_lun
writeu, unit, wavelength, Zdist
writeu, unit, Xnum, Ynum
writeu, unit, ccdsp, Ax
close, unit
free_lun, unit

end

```

2. Program to reconstruct the digital hologram generated by the program above via Fresnel approximation algorithm

```

PRO FT_RECON_pbp
; Numerical reconstruction by Fresnel transformation with the simulations of
point-to-point addition
; Reconstruction of the real image with using lens function

; Read the set-up parameters
wavelength = 0.0
Zdist = 0.0
Xnum = 0
Ynum = 0
ccdsp = 0.0
Angle = 0.0
file = dialog_pickfile(/read, filter = '*.dat')
datafilename = file
openr, unit, datafilename, /get_lun
readu, unit, wavelength, Zdist
readu, unit, Xnum, Ynum
readu, unit, ccdsp, Angle
close, unit
free_lun, unit
Xlength = (Xnum - 1) * ccdsp
Ylength = (Ynum - 1) * ccdsp

; Read the digital hologram
hologram = fltarr(Xnum, Ynum)
file = strmid(file, 0, strlen(file)-4) + '.hol'
holofilename = file
openr, unit, holofilename, /get_lun
readu, unit, hologram
close, unit
free_lun, unit

; Initiate the data arrays
ccdPx = fltarr(Xnum)
ccdPy = fltarr(Ynum)
PROP_REAL = FLTARR(XNUM, YNUM)
PROP_IMAG = FLTARR(XNUM, YNUM)
THETA2_REAL = FLTARR(XNUM, YNUM)
THETA2_IMAG = FLTARR(XNUM, YNUM)
LINE_REAL = FLTARR(XNUM)
LINE_IMAG = FLTARR(XNUM)
ref_real = fltarr(Xnum, Ynum)
ref_imag = fltarr(Xnum, Ynum)

; Suppression of the DC term
ave_holo = total(hologram) / (float(Xnum) * float(Ynum))
hologram = hologram - ave_holo

; Create the coordinate system of CCD array
ccdPx[0] = -1. * Xlength / 2.
for i = 1, Xnum - 1 do begin
    ccdPx[i] = ccdPx[i-1] + ccdsp
endfor
ccdPy[0] = -1. * Ylength / 2.
for i = 1, Ynum - 1 do begin
    ccdPy[i] = ccdPy[i-1] + ccdsp
endfor

```

```

; Generate plane reference wave for reconstruction
ref_amp = 1.0
wavenum = 2. * !pi / wavelength
for x=0, Xnum-1 do begin
    phase_angle = wavenum * sin(Angle) * ccdPx[x]
    line_real[x] = ref_amp * cos(phase_angle)
    line_imag[x] = ref_amp * sin(phase_angle)
endfor
for y=0, Ynum-1 do begin
    ref_real[:,y] = line_real
    ref_imag[:,y] = line_imag
endfor

; Numerical reconstruction by Fresnel approximation
var1_real = fltarr(Xnum, Ynum)
var1_imag = fltarr(Xnum, Ynum)
factor1 = -1.0 * !pi / (wavelength * Zdist)
for x=0, Xnum - 1 do begin
    Xeuclid = ccdPx[x] ^ 2.
    for y=0, Ynum - 1 do begin
        Yeuclid = ccdPy[y] ^ 2.
        euclid = Xeuclid + Yeuclid
        var1_real[x, y] = cos(factor1 * euclid)
        var1_imag[x, y] = sin(factor1 * euclid)
    endfor
endfor
var2_real = hologram * (ref_real * var1_real - ref_imag * var1_imag)
var2_imag = hologram * (ref_real * var1_imag + ref_imag * var1_real)
c = complex(var2_real, var2_imag)
f = fft(c, /inverse)
f = shift(f, Xnum/2, Ynum/2)
c_real = real_part(f)
c_imag = imaginary(f)
factor2 = -1.0 * !pi * wavelength * Zdist
var3_real = fltarr(Xnum, Ynum)
var3_imag = fltarr(Xnum, Ynum)
for x=0, Xnum-1 do begin
    Xdist = ((x-(Xnum-1)/2) / (Xlength + ccdsp)) ^ 2.
    for y=0, Ynum-1 do begin
        Ydist = ((y-(Ynum-1)/2) / (Ylength + ccdsp)) ^ 2.
        dist = Xdist + Ydist
        var3_real[x,y] = cos(factor2 * dist)
        var3_imag[x,y] = sin(factor2 * dist)
    endfor
endfor
factor3 = -1.0 * wavenum * Zdist
var4_real = (sin(factor3) * var3_real - cos(factor3) * var3_imag) / (wavelength * Zdist)
var4_imag = (sin(factor3) * var3_imag + cos(factor3) * var3_real) / (wavelength * Zdist)
ccd_real = var4_real * c_real - var4_imag * c_imag
ccd_imag = var4_real * c_imag + var4_imag * c_real

; Calculate the reconstructed intensity image on CCD and display it
ccd = ccd_real ^ 2.0 + ccd_imag ^ 2.0
phase = atan(ccd_imag, ccd_real)
window, 0, xsize = 512, ysize = 512, title = 'Reconstructed intensity image'
tvsc1, congrid(ccd, 512, 512)
window, 1, xsize = 512, ysize = 512, title = 'Reconstructed phase image'
tvsc1, congrid(phase, 512, 512)

end

```


3. Program to reconstruct the digital hologram genera by convolution approach

```

PRO CV1_RECON

; Reconstruction by convolution approach with 3 FFTs

; Read the set-up parameters
Wavelength = 0.0
Zdist = 0.0
Xnum = 0
Ynum = 0
ccdsp = 0.0
Angle = 0.0
file = dialog_pickfile(/read, filter = '*.dat')
datafilename = file
openr, unit, datafilename, /get_lun
readu, unit, Wavelength, Zdist
readu, unit, Xnum, Ynum
readu, unit, ccdsp, Angle
close, unit
free_lun, unit
Xlength = (Xnum - 1) * ccdsp
Ylength = (Ynum - 1) * ccdsp

; Read the digital hologram
hologram = fltarr(Xnum, Ynum)
file = strmid(file, 0, strlen(file)-4) + '.hol'
holofilename = file
openr, unit, holofilename, /get_lun
readu, unit, hologram
close, unit
free_lun, unit

; Initiate the data arrays
ccd_x = fltarr(Xnum)
ccd_y = fltarr(Ynum)
prop_real = fltarr(Xnum, Ynum)
prop_imag = fltarr(Xnum, Ynum)
theta2_real = fltarr(Xnum, Ynum)
theta2_imag = fltarr(Xnum, Ynum)
line_real = fltarr(Xnum)
line_imag = fltarr(Xnum)
ref_real = fltarr(Xnum, Ynum)
ref_imag = fltarr(Xnum, Ynum)

; Suppression of the DC term
ave_holo = total(hologram) / (float(Xnum) * float(Ynum))
hologram = hologram - ave_holo

; Create the coordinate system of CCD array
ccd_x[0] = -1. * Xlength / 2.
for i = 1, Xnum - 1 do begin
    ccd_x[i] = ccd_x[i-1] + ccdsp
endfor
ccd_y[0] = -1. * Ylength / 2.
for i = 1, Ynum - 1 do begin
    ccd_y[i] = ccd_y[i-1] + ccdsp
endfor

```

```

; Generate plane reference wave for reconstruction
ref_amp = 1.0
wavenum = 2. * !pi / Wavelength
for x=0, Xnum-1 do begin
    phase_angle = -1. * wavenum * sin(Angle) * ccd_x[x]
    line_real[x] = ref_amp * cos(phase_angle)
    line_imag[x] = ref_amp * sin(phase_angle)
endfor
for y=0, Ynum-1 do begin
    ref_real[:,y] = line_real
    ref_imag[:,y] = line_imag
endfor

; Numerical reconstruction by convolution approach
var1_real = hologram * ref_real
var1_imag = hologram * ref_imag
var1 = complex(var1_real, var1_imag)
c = fft(var1)
var1_real = float(c)
var1_imag = imaginary(c)
var2_real = fltarr(Xnum, Ynum)
var2_imag = fltarr(Xnum, Ynum)
Z_dist = Zdist ^ 2.0
for x=0, Xnum - 1 do begin
    X_dist = ccd_x[x] ^ 2.
    for y=0, Ynum - 1 do begin
        Y_dist = ccd_y[y] ^ 2.
        dist = sqrt(X_dist + Y_dist + Z_dist)
        var2_real[x, y] = sin(wavenum * dist) / (dist * Wavelength)
        var2_imag[x, y] = cos(wavenum * dist) / (dist * Wavelength)
    endfor
endfor
var2 = complex(var2_real, var2_imag)
c = fft(var2)
var2_real = float(c)
var2_imag = imaginary(c)
var3_real = var1_real * var2_real - var1_imag * var2_imag
var3_imag = var1_real * var2_imag + var1_imag * var2_real
var3 = complex(var3_real, var3_imag)
c = fft(var3, /inverse)
c = shift(c, Xnum/2, Ynum/2)
recon_real = float(c)
recon_imag = imaginary(c)

; Calculate the reconstructed intensity image on CCD and display it
ccd = recon_real ^ 2.0 + recon_imag ^ 2.0
phase = atan(recon_imag, recon_real)
window, 0, xsize = 512, ysize = 512, title = 'Reconstructed intensity image'
tvsc1, ccd
window, 1, xsize = 512, ysize = 512, title = 'Reconstructed phase image'
tvsc1, phase

end

```

4. Program for the shape measurement of a ball bearing by two-source contouring

; Two-source contouring for a ball bearing. The digital holograms used in this case were captured in 4-scan mode.

```

; Function for curve fitting dependent on the shape of the object under study
pro gfunct, X, A, F, pder
    bx = sqrt(A[2]^2-(X-A[0])^2)
    F = bx + A[1]
    if n_params() ge 4 then $
        pder = [(X-A[0])/bx], [replicate(1.0, n_elements(X))], [A[2]/bx]]
end

; Parameters of the holographic system
Wavelength = 0.0006328 ; wavelength used in the contouring
Zdist = -355. ; the distance between the object plane and CCD plane
ccdsp = 0.003225 ; pixel spacing in the CCD
Xnum = 0 ; pixel number in x dimension
Ynum = 0 ; pixel number in y dimension
Angle = 0. * !pi / 180 ; the angle of the reference wave
ref_amp = 1.0 ; the amplitude of the plane reference wave
wavenum = 2 * !PI / wavelength ; wave number
illu_angle = -32. * !pi / 180 ; illumination angle of the object wave
in the first exposure
del_angle = 0.01 * !pi / 180 ; the change of the illumination angle

; Read the holograms captured with different illumination angles
file1 = 'D:\RSI\IDL61\experiment results\2008-12-15\holo_081215_001.BMP'
hologram1_3d = read_image(file1)
b = size(hologram1_3d)
Xnum = b[3]
Ynum = b[3]
ccdXlength = (Xnum - 1) * ccdsp ; CCD length in x dimension
ccdYlength = (Ynum - 1) * ccdsp ; CCD length in y dimension
file2 = strmid(file1, 0, strlen(file1)-6) + '02.BMP'
hologram2_3d = read_image(file2)
hologram1 = fltarr(Xnum, Ynum)
hologram2 = fltarr(Xnum, Ynum)
x_dis = (b[2]-b[3])/2
for x = 0, Xnum-1 do begin
    for y = 0, Ynum-1 do begin
        hologram1[x,y] = hologram1_3d[0,x+x_dis,y]
        hologram2[x,y] = hologram2_3d[0,x+x_dis,y]
    endfor
endfor

; Suppression of the zero-order term
ave_holo1 = total(hologram1) / n_elements(hologram1)
hologram1 = hologram1 - ave_holo1
ave_holo2 = total(hologram2) / n_elements(hologram2)
hologram2 = hologram2 - ave_holo2

; Initiate the data arrays
ccdPx = fltarr(Xnum, Ynum)
ccdPy = fltarr(Xnum, Ynum)
orderXX = fltarr(Xnum, Ynum)
orderYY = fltarr(Xnum, Ynum)
objPx = fltarr(Xnum, Ynum)
objPy = fltarr(Xnum, Ynum)
plX = fltarr(Xnum)
plY = fltarr(Ynum)

```

```

; Create the coordinate system of the CCD plane
plX[0] = -1.0 * ccdXlength / 2.0
for i = 1, Xnum - 1 do begin
    plX[i] = plX[i-1] + ccdsp
endfor
for i = 0, Ynum - 1 do begin
    ccdPx[*, i] = plX
endfor
plY[0] = -1.0 * ccdYlength / 2.0
for i = 1, Ynum - 1 do begin
    plY[i] = plY[i-1] + ccdsp
endfor
for i = 0, Xnum - 1 do begin
    ccdPy[i, *] = plY
endfor

; Create the coordinate system of the object plane
objXsp = abs(wavelength * Zdist / (Xnum * ccdsp))
objYsp = abs(wavelength * Zdist / (Ynum * ccdsp))
objXlength = (Xnum - 1) * objXsp
objYlength = (Ynum - 1) * objYsp
plX[0] = -1.0 * objXlength / 2.0
for i = 1, Xnum - 1 do begin
    plX[i] = plX[i-1] + objXsp
endfor
for i = 0, Ynum - 1 do begin
    objPx[*, i] = plX
endfor
plY[0] = -1.0 * objYlength / 2.0
for i = 1, Ynum - 1 do begin
    plY[i] = plY[i-1] + objYsp
endfor
for i = 0, Xnum - 1 do begin
    objPy[i, *] = plY
endfor

; Generate the reference wave
ref_phase = -2 * !pi * ccdPx * sin(Angle) / Wavelength
ref_real = ref_amp * cos(ref_phase)
ref_imag = ref_amp * sin(ref_phase)

; Numerical reconstruction of these two holograms by Fresnel approximation
orderX = fltarr(Xnum)
orderY = fltarr(Ynum)
orderX[0] = (1 - Xnum) / 2.
for i = 1, Xnum - 1 do begin
    orderX[i] = orderX[i-1] + 1
endfor
for i = 0, Ynum - 1 do begin
    orderXX[*, i] = orderX
endfor
orderY[0] = (1 - Ynum) / 2.
for i = 1, Ynum - 1 do begin
    orderY[i] = orderY[i-1] + 1
endfor
for i = 0, Xnum - 1 do begin
    orderYY[i, *] = orderY
endfor
para1 = -1.0 * !pi / (Wavelength * Zdist)
para2 = ccdPx ^ 2.0 + ccdPy ^ 2.0
para3_real = cos(para1 * para2)

```

```

para3_imag = sin(para1 * para2)
para41_real = hologram1 * (ref_real * para3_real - ref_imag * para3_imag)
para41_imag = hologram1 * (ref_real * para3_imag + ref_imag * para3_real)
para42_real = hologram2 * (ref_real * para3_real - ref_imag * para3_imag)
para42_imag = hologram2 * (ref_real * para3_imag + ref_imag * para3_real)
c1 = complex(para41_real, para41_imag)
c2 = complex(para42_real, para42_imag)
f1 = fft(c1, /inverse)
f2 = fft(c2, /inverse)
f1 = shift(f1, Xnum/2, Ynum/2)
f2 = shift(f2, Xnum/2, Ynum/2)
c1_real = float(f1)
c1_imag = imaginary(f1)
c2_real = float(f2)
c2_imag = imaginary(f2)
para5 = -1.0 * !pi * Wavelength * Zdist
para6 = (orderXX / (ccdXlength + ccdsp)) ^ 2.0 + (orderYY / (ccdYlength + ccdsp))
^ 2.0
para7_real = cos(para5 * para6)
para7_imag = sin(para5 * para6)
para81_real = para7_real * c1_real - para7_imag * c1_imag
para81_imag = para7_real * c1_imag + para7_imag * c1_real
para82_real = para7_real * c2_real - para7_imag * c2_imag
para82_imag = para7_real * c2_imag + para7_imag * c2_real
Wavenum = 2 * !pi / Wavelength
para9_real = sin(Wavenum * Zdist) / (Wavelength * Zdist)
para9_imag = cos(Wavenum * Zdist) / (Wavelength * Zdist)
para01_real = para81_real * para9_real - para81_imag * para9_imag
para01_imag = para81_real * para9_imag + para81_imag * para9_real
para02_real = para82_real * para9_real - para82_imag * para9_imag
para02_imag = para82_real * para9_imag + para82_imag * para9_real

; Obtain and display the phase-difference image
phase = fltarr(Xnum, Ynum)
phase1 = atan(para01_imag, para01_real)
phase2 = atan(para02_imag, para02_real)
for x=0, Xnum-1 do begin
    for y=0, Ynum-1 do begin
        if (phase2[x,y] lt phase1[x,y]) then phase[x,y] = phase2[x,y] -
phase1[x,y] + 2 * !pi $
        else phase[x,y] = phase2[x,y] - phase1[x,y]
    endfor
endfor
window, 0, xsize = 512, ysize = 512, title = 'direct subtracted phase'
tvsc1, congrid(phase, 512, 512)

; Choose the region of interest
cursor, left, down, /down, /device
cursor, right, up, /up, /device
left = left * 4
right = right * 4
up = up * 4
down = down * 4
Xdim = right - left + 1
Ydim = down - up + 1

; Remove the linear factor caused by tilted illumination
incl_i_factor = 2 * wavenum * sin(del_angle/2) * cos(illu_angle+del_angle/2)
incl_i_phase = incl_i_factor * objPx[left:right, up:down]
incl_i_phase = atan(sin(incl_i_phase), cos(incl_i_phase))+!pi
nor_phase = fltarr(Xdim, Ydim)

```

```

r_phase = phase[left:right, up:down]
for x=0, Xdim-1 do begin
    for y=0, Ydim-1 do begin
        if (incl_i_phase[x,y] lt r_phase[x,y]) then nor_phase[x,y] =
incl_i_phase[x,y] - r_phase[x,y] + 2 * !pi $
        else nor_phase[x,y] = incl_i_phase[x,y] - r_phase[x,y]
    endfor
endfor
window, 1, xsize = Xdim, ysize = Ydim
tvsc1, nor_phase ; display the phase-difference image after the removal of
the linear factor

; Unwrap the phase-difference image by calling a user-specified procedure and
display the unwrapped phase
unwrapped = unwrapimage2(nor_phase, 4, Xdim, Ydim, 3)
window, 2, xsize = Xdim, ysize = Ydim
tvsc1, unwrapped

; Get the height distribution of the surface
height = unwrapped / (wavenum * 2 * sin(del_angle/2) *
sin(illu_angle+del_angle/2))
height = smooth(height, 10, /edge_truncate)

; Generate the mask for the object under study
cursor, cx, cy, /device
area_r = 2001
height_r = fltarr(Xdim, Ydim)
for i=0, Xdim-1 do begin
    for j=0, Ydim-1 do begin
        if (long(i-cx))^2+(long(j-cy))^2 gt area_r^2 then height_r[i,j] = 0 $
        else height_r[i,j] = height[i,j] - height[cx-area_r,cy]
    endfor
endfor

; Display the surface of the object
x_cor = plX[left:right] - plX[left]
y_cor = plY[up:down] - plY[up]
window, 3
shade_surf, height_r, x_cor, y_cor, ax=45, az=30,
zrange=[min(x_cor),max(x_cor)]

; Find the maximum height and its position
print, 'max_h = ', max(height_r,i)
mx = i mod Xdim
print, mx
my = i / Xdim
print, my

; Display the cross sections along the x and y axes from the position of the
maximum height
window, 4, xsize = 512, ysize = 256, title = 'profile x'
plot, x_cor, height_r[*,my], /isotropic
window, 5, xsize = 512, ysize = 256, title = 'profile y'
plot, y_cor, height_r[mx,*], /isotropic

; Curve fitting for both cross sections displayed above. The fitting parameters
vary depending upon the holograms used in the contouring.
i = 0

```

```

while (height_r[i,my] le 2) and (i lt Xdim) do i = i + 1
mxl = i
i = i + 100
while (height_r[i,my] gt 2.5) and (i lt Xdim) do i = i + 1
mxr = i

j = 0
while (height_r[mx,j] le 2) and (j lt Ydim) do j = j + 1
myl = j
j = j + 100
while (height_r[mx,j] gt 3) and (j lt Ydim) do j = j + 1
myr = j

X1 = x_cor[mxl:mxr]
Y1 = height_r[mxl:mxr,my]
weights = replicate(1.0, n_elements(X1))
A1 = [9., -1., 8.]
yfit = curvefit(X1,Y1,weights,A1,sigma,function_name='gfunct',yerror=b1)
print, 'Function parameters: ', A1, b1

X2 = y_cor[myl:myr]
Y2 = height_r[mx,myl:myr]
weights = replicate(1.0,n_elements(X2))
A2 = [9.75, -1., 8.]
yfit = curvefit(X2,Y2,weights,A2,sigma,function_name='gfunct',yerror=b2)
print, 'Function parameters: ', A2, b2

end

```